Théorie du risque en santé publique: Application aux infections en milieu hospitalier

Mathieu Emily

Théorie du risque en santé publique: Application aux infections en milieu hospitalier

Mathieu Emily

Laboratoire de Statistique, Université Haute Bretagne, Place du Recteur H. Le Moal, CS 24307, 35043 Rennes Cedex

Résumé

Nous introduisons une nouvelle approche pour traiter le problème des infections contractées lors d'un séjour hospitalier. En nous inspirant de la théorie du risque en actuariat, nous avons défini un coefficient d'ajustement permettant de quantifier le risque de développer une infection dans un service hospitalier. Ce coefficient caractérise une distribution exponentielle de la fonction de survie et est estimé comme solution d'une équation de Lundberg. La robustesse des estimateurs a été évaluée sur données simulées. De plus, notre approche a été appliquée à des données du programme de médicalisation des systèmes d'informations (PMSI) concernant le risque d'embolie pulmonaire suite à une prothèse de hanche dans la région Rhône-Alpes.

Abstract

We introduce a new approach to hospital-acquired disease risk assessment from public health databases. In a spirit similar to actuarial theory, we define an adjustment coefficient that can quantify the risk associated with hospital department. The adjustment coefficient characterizes the tail of the distribution of patient length of stay and we show that this coefficient is a solution of Lundberg-like equation. Using simulations, we provide evidence of the robustness of the approximation. In addition, we illustrate the relevance of this approach on hip replacement procedure, after which patient may contract pulmonary embolism.

Mot-clés: Risque en santé publique, Coefficient d'ajustement, Infection, Durée de séjour hospitalier.

1 Risque en santé publique

Depuis plusieurs années, les erreurs médicales et les complications liées à un séjour hospitalier sont devenues une préoccupation importante dans le domaine de la santé publique (Institute of Medicine, 2001). Dans de nombreux pays, des systèmes de surveillance spécialisés ont vu le jour et ont permis de mettre en évidence certaines améliorations possibles du système de santé. Parmi elles, l'amélioration du contrôle du risque d'infections contractées pendant un séjour hospitalier, et notamment les infections nosocomiales, nécessite le développement d'outils statistiques appropriés (Iezzoni, 2003).

Pour la plupart des infections, la durée de séjour en milieu hospitalier constitue un des principal facteurs de risque. La prise en compte explicite du temps de séjour dans la modélisation du risque individuel d'infection permet un parallèle avec la théorie du risque appliquée à l'actuariat. Pour cet article, nous nous proposons d'adapter des outils d'actuariat, comme la modélisation du risque de ruine d'une compagnie d'assurance, au risque d'infection en milieu hospitalier. Notre modèle se focalise sur l'expression de la fonction survie, $\Psi(t) = P(T > t)$, où T modélise le temps avant la prochaine infection.

Dans un premier temps, nous proposerons une approximation de Cramer-Lundberg pour résoudre l'équation de renouvellement régissant le comportement de la fonction de survie. Dans ce cadre, nous développerons notamment un estimateur non-paramétrique du coefficient d'ajustement de Lundberg. Notre modèle sera par la suite évalué sur données simulées. Afin d'illustrer notre approche, nous nous sommes intéressés à l'occurence d'une embolie pulmonaire suite à une pose de prothèse de hanche dans les hôpitaux de la région Rhône-Alpes.

2 Modélisation du risque d'infection

Nous nous intéressons à la distribution du nombre de jours d'hospitalisation avant l'occurence d'une infection, dont une série de k observations sera notée par la suite $T_1, T_2, \ldots, T_k$. Entre deux infections, n patients se sont succédés dans le service. Les durées de séjour de ces patients, notées $X_1, X_2, \ldots, X_n$, sont supposées indépendantes et identiquement distribuées, de densité inconnue $f(x)$ et de fonction de répartition $F(x)$. A chaque durée de séjour, nous associons une variable binaire, notée $Z_1, Z_2, \ldots$, indiquant un événement d'infection (ou non) au cours du séjour. La probabilité conditionnelle d'un événement d'infection sachant la durée de séjour vaut x, $p(x) = P(Z_i = 1 | X_i = x)$, est inconnue.

Notons à présent $I_0 = 0$ et, pour $j \geq 1$, $I_j = \min \{i > I_{j-1}, Z_i = 1\}$, correspondant au patient contractant la $j^{\text{ème}}$ infection. Nous en déduisons donc que:

$$T_j = \sum_{i=I_{j-1+1}}^{I_j} X_i$$

Sous les hypothèses précédentes, les variables T_i sont $i.i.d$ et posons T une variable aléatoire suivant la même loi que les T_i. Dans ce contexte, la fonction de survie se définie par :

$$\Psi(t) = P(T > t), \quad t > 0.$$

Grâce à la théorie du renouvellement (Resnick, 1992 et Ross, 1996), nous obtenons, pour la fonction de survie, l'équation de renouvellement suivante :

$$\Psi(t) = 1 - F(t) + \int_0^t \Psi(t-x)(1-p(x))f(x)dx, \quad t > 0 \tag{1}$$

Approximation de Cramer-Lundberg. Dans la théorie du risque en actuariat (Cramer, 1930), la probabilité de ruine d'une compagnie d'assurance ϕ peut être approchée par:

$$\phi(u) \approx 1 - C_R e^{-Ru}$$

Par analogie, nous allons approcher la probabilité de non-occurence d'une infection pendant une période t, c'est à dire la fonction de survie définie à l'équation 1, par:

$$\Psi(t) = C_R e^{-Rt}$$

où les coefficients C_R et R vont être estimés. R est appelé le coefficient d'ajustement et se trouve être la solution de l'équation de Lundberg (Grandell, 1991):

$$\int_0^\infty e^{Rx}(1-p(x))f(x)dx = 1 \tag{2}$$

Par la théorie du renouvellement, nous obtenons également que :

$$C_R = \frac{\int_0^\infty e^{Rx}(1-F(x))dx}{\int_0^\infty xe^{Rx}(1-p(x))f(x)dx} \tag{3}$$

Estimation de R et C_R. Les équations 2 et 3 n'étant pas calculables, leur utilisation pour l'inférence des paramètres R et C_R est impossible. Cependant, en conditionnant à l'absence d'infection, nous pouvons définir des nouvelles durées de séjour, notées Y_i, telles que :

$$P(Y_i \leq s) = P(X_i \leq s | Z_i = 0) \quad s \geq 0$$

La variable aléatoire, correspondante à la durée de séjour conditionnelle à l'absence d'infection, sera notée $Y = (X|Z = 0)$.

L'utilisation de la formule de Bayes, combinée à l'équation 2, nous donne la formulation suivante :

$$E[e^{RY}] = \frac{1}{P(\text{Absence d'infection})}$$

L'expression ci-dessus nous permet de définir un estimateur $\hat{R}$ de R de la façon suivante:

$$\frac{1}{n}\sum_{i=1}^{n} e^{\hat{R}y_i} = \frac{1}{1-\bar{p}}, \tag{4}$$

où, les y_i correspondent aux durées de séjour pour lesquelles aucune infection n'a été détectée, n est le nombre total d'observations et $\bar{p}$ une estimation du risque individuel.

De façon analogue, nous obtenons que :

$$C_R = \frac{\frac{E[e^{RX}]-1}{R}}{P(\text{Absence d'infection}) \times E[Ye^{RY}]} \tag{5}$$

L'approximation des espérances par la moyenne empirique dans l'expression ci-dessus nous fournit un estimateur du coefficient C_R.

La section suivante est dédiée à l'étude des performances des estimateurs proposés par les équations 4 et 5.

3 Résultats par simulation et étude de cas

Etude par simulations. Afin d'étudier le comportement des estimateurs des coefficients R et C_R, définis à la section précédente (équations 4 et 5), nous proposons de simuler d'une part les durées de séjours selon une loi exponentielle de paramètre λ, et d'autre part les probabilités conditionnelles d'infection $p(x)$ selon 10 modèles. Ces modèles, allant de mélanges de lois exponentielles à des lois uniformes en passant par des lois Gamma et de Weibull (voir Tableau 1), représentent un large spectre de distributions réalistes.

Afin d'évaluer les performances des estimateurs par Monté-Carlo, 1000 durées de séjour, associées à l'occurence d'un événement d'infection, ont été simulées pour chaque simulation. Pour chacun des 10 modèles de probabilités conditionnelles d'infection, 10000 simulations ont été effectuées. Le tableau 1 résume les résultats obtenus, par l'intermédiaire de la moyenne et de l'écart-type des estimateurs $\hat{R}$ et $\hat{C_R}$.

Nous pouvons constater que les estimateurs sont sans biais avec un écart-type faible. Seuls les modèles de loi de Weibull(1/3,1) et de loi Gamma(3,1/2) présentent des résultats peu convaincants.

Analyse des données d'embolie pulmonaire suite à une prothèse de hanche. Afin de valider cette approche sur un jeu de données réelles, nous nous sommes intéressés à des patients opérés pour une prothèse de hanche au cours de l'année 2006 dans 20 hôpitaux de la région Rhône-Alpes (données extraites du programme de médicalisation des sytèmes d'information ou PMSI). Suite à cette opération, certains patients ont développé une embolie pulmonaire. Pour chacun des 20 hôpitaux étudiés, nous avons estimé le coefficient d'ajustement ($\hat{R}$) grâce à l'équation 4 :

$$\frac{1}{n}\sum_{i=1}^{n} e^{\hat{R}y_i} = \frac{1}{1-\bar{p}},$$

Lorsque $\hat{R} \approx 0$, caractérisant ainsi une embolie pulmonaire comme un événement rare, l'équation précédente peut se simplifier par $\hat{R} \approx \bar{p}/(\bar{y}(1-\bar{p}))$, où $\bar{y}$ est la moyenne

Risk Model	R	$\bar{\hat{R}}$	$\sigma(\hat{R})$	C_R	$\bar{\hat{C}}_R$	$\sigma(\hat{C}_R)$	λ
Gamma(3/2,2)	0.2031	0.20	0.02	1.1529	1.15	0.03	2
Gamma(1/2,5)	0.7403	0.74	0.05	1.3397	1.35	0.10	2
Gamma(3,1/2)	0.5324	0.53	0.03	2.8361	2.86	0.53	1
Weibull(1/3,1)	0.7076	0.73	0.06	1.5945	2.13	3.21	1
Weibull(2,5)	0.0750	0.07	0.01	1.1533	1.15	0.03	1
Weibull(3,2)	0.3326	0.33	0.02	2.0189	2.02	0.15	1
Unif(0,1)	1.1268	1.13	0.11	1.1454	1.15	0.03	10
Exp(0.1)	0.0999	0.10	0.01	1.1111	1.11	0.02	1
0.2 Exp(0.02) +0.8Exp(0.09)	0.0751	0.07	0.01	1.0804	1.08	0.02	1
0.3 Exp(0.01) +0.4Exp(0.02) +0.3Exp(1)	0.1777	0.18	0.01	1.1185	1.12	0.03	1

Table 1: Performances des estimateurs $\hat{R}$ et $\hat{C_R}$. La colonne R (resp. C_R) reporte les vrais valeurs de R (resp. C_R) sous le modèle considéré. $h\bar{at}R$ et $\sigma(\hat{R})$ (resp. $h\bar{at}C_R$ et $\sigma(\hat{C_R})$) correspondent aux estimations Monté-Carlo de l'espérance et de l'écart-type de $\hat{R}$ (resp. $\hat{C_R}$)

des y_i. La figure 1 représente l'inverse des durées moyennes de séjours sans embolie ($\bar{y}$) en fonction des valeurs estimées du coefficients d'ajustement pour les 20 hôpitaux. Les données ont été normalisées par le coefficient de régression du risque et de l'inverse de durée de séjour.

Nous pouvons remarquer que deux hôpitaux s'écartent de la tendance globale (hôpital 3 et hôpital 13). Ces résultats ont été confirmés par une analyse complémentaire, pour laquelle un modèle de régression logistique a été utilisé pour expliquer une embolie pulmonaire en fonction des numéros des hôpitaux.

4 Discussion et conclusion

Dans cet article, nous proposons un modèle, inspiré de la théorie du risque en actuariat, permettant d'évaluer le risque de développer une infection suite à un séjour dans un service hospitalier. Par l'intermédiaire d'un coefficient d'ajustement, R, et d'un coefficient d'échelle, C_R, la fonction de survie est modélisée comme une fonction exponentielle $\Psi(t) \approx C_R e^{-Rt}$.

En s'appuyant sur la théorie des processus de renouvellement, nous avons proposé des estimateurs des coefficients R et C_R permettant ainsi une évaluation de la probabilité de développer une infection. Les performances de ces estimateurs ont été évalués par simulations et ont montré des résultats convaincants, quant au biais et à la variance,

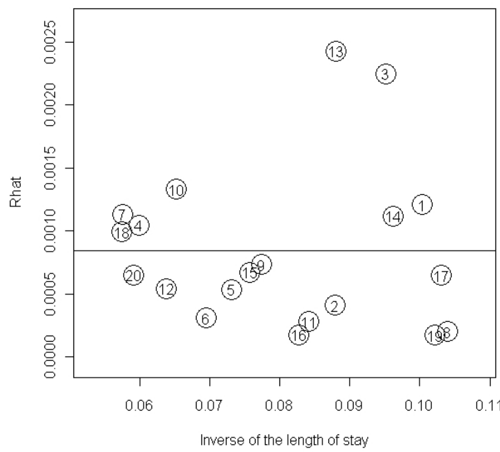

Figure 1: Coefficient d'ajustement en fonction de l'inverse des durées de séjour. La droite correspond à la regression du risque et de l'inverse des durées de séjour.

sur 10 modèles spécifiques. La procédure développée a ensuite été appliquée à un jeu de données, résumant le développement d'une embolie pulmonaire suite à une prothèse de hanche, dans 20 établissements de la région Rhône-Alpes. Les résultats ont permis de mettre en évidence deux établissements pour lesquels le taux d'embolie pulmonaire semble anormal.

L'ensemble de ces résultats montre que l'élaboration d'outils statistiques adaptés permet d'obtenir des informations précises sur le fonctionnement de certains établissement hospitaliers. Développer de tels modèles peut avoir des répercussions sur la politique de santé publique décidée à différentes échelles: à l'échelle du service, à l'échelle de l'établissement voire même à l'échelle nationale. De plus, en mettant en parallèle différents domaines d'applications, comme l'actuariat et la santé publique ici, ces modèles statistiques trouveront des applications plus larges.

Bibliographie

[1] Institue of Medicine (2001). Crossing the Quality Chasm: A New Health System for the 21st Century. Committee of Quality of Care in America, Washington DC: National Academy Press.

[2] Iezzoni, L. I. (2003) *Adjustment for measuring health care outcomes*, 3rd ed. Chicago: Academy Health and Health Administration Press.

[3] Resnick, S. (1992) *Adventures in Stocahstic Processes*, Boston, Birkhaüser.

[4] Ross, S. M. (1996) *Stochastic Processes*, 2nd ed. New York, Wiley.

[5] Cramer, H. (1930) *On the mathematical theory of risks*, Skandia Jubilee Volume Stockholm.

[6] Grandell, J. (1991) *Aspects of Risk Theory*, Springer-Verlag, New York.