

Analyse statistique des risques actuariels

Vraisemblance

On peut voir la fonction de densité $f(x; \theta)$ comme étant une fonction du paramètre inconnu θ avec x fixé; ceci est la fonction de vraisemblance $\mathcal{L}(\theta; x)$.

Qualité de l'estimateur

La première section traite de «**estimateurs ponctuels**». C'est-à-dire, on produit une seule valeur comme notre meilleur essai pour déterminer la valeur de la population inconnue. Intrinsèquement, on ne s'attend pas à ce que cette valeur (même si c'en est une bonne) soit la vraie valeur exacte.

Une hypothèse plus utile à des fins d'interprétation est plutôt un **estimateur par intervalle**; au lieu d'une seule valeur, il retourne un intervalle de valeurs plausibles qui peuvent toutes être la vraie valeur. Le type principal d'*estimateur par intervalle* est l'*intervalle de confiance* traité dans la deuxième sous-section.

Estimation ponctuelle

Lorsque nous avons un estimateur $\hat{\theta}$ pour un paramètre inconnu θ on espère que, **en moyenne**, ses erreurs de prévision vont s'annuler. On peut alors trouver $E\hat{\theta}|\theta$, soit l'espérance de l'estimateur sachant que θ est la vraie valeur du paramètre. Par la suite, on calcule son **biais** $B(\hat{\theta})$ dans la prévision du paramètre :

Biais d'un estimateur

$$B(\hat{\theta}) = E\hat{\theta}|\theta - \theta$$

Cependant, le biais n'indique pas la variabilité de l'estimateur $\hat{\theta}$ dans sa prévision. Nous définissons alors la **borne inférieure Cramèr-Rao** de la variance de l'estimateur $\text{Var}(\hat{\theta})$ avec l'**information de Fisher** $I(\theta)$:

Borne inférieure Cramèr-Rao

$$\text{Var}(\hat{\theta}) \geq \frac{1}{nE \left[\left(\frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 \right]}, \quad \text{où } I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 \right]$$

Par la suite, nous pouvons définir l'**efficacité** d'un estimateur comme étant le ratio de la borne Cramèr-Rao sur la variance de l'estimateur :

Efficacité d'un estimateur

$$\text{eff}(\theta) = \frac{\text{Var}(\hat{\theta})^{\text{Rao}}}{\text{Var}(\hat{\theta})} = \frac{1}{n\text{Var}(\hat{\theta})E \left[\left(\frac{\partial}{\partial \theta} \ln f(x; \theta) \right)^2 \right]}$$

Si ce ratio est de 1 $\text{eff}(\theta) = 1$, alors l'estimateur est **efficace**.

Nous pouvons aussi généraliser cette formulation pour obtenir l'efficacité relative d'un estimateur contre un autre :

Efficacité relative

Soit les estimateurs non-biaisés $\hat{\theta}_n$ et $\tilde{\theta}_n$, l'efficacité de $\hat{\theta}_n$ relatif à $\tilde{\theta}_n$ est :

$$\text{eff}(\hat{\theta}_n, \tilde{\theta}_n) = \frac{\text{Var}(\hat{\theta}_n)}{\text{Var}(\tilde{\theta}_n)}$$

Si $\text{eff}(\hat{\theta}_n, \tilde{\theta}_n) < 1$ alors $\hat{\theta}_n$ est mieux et vice-versa.

Nous pouvons également évaluer si un estimateur est cohérent, ou converge, avec des très grands échantillons; un estimateur $\hat{\theta}$ est **consistant** (*angl.*) si la probabilité qu'il diffère de la vraie valeur du paramètre θ par une erreur ϵ près de 0 tends vers 0 alors que la taille de l'échantillon n tends vers l'infini :

Convergence (consistency) d'un estimateur

$$\lim_{n \rightarrow \infty} \Pr(|\hat{\theta} - \theta| > \epsilon) = 0, \quad \epsilon > 0$$

Ce critère peut être rencontré lorsque l'estimateur $\hat{\theta}$ est **asymptotiquement sans biais** et que la **variance de l'estimateur** tend vers 0. En premier temps, on définit ce qu'est un estimateur asymptotiquement sans biais :

Estimateur asymptotiquement sans biais

$$\lim_{n \rightarrow \infty} B(\hat{\theta}) = 0$$

Donc si $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}) = 0$ et que $\lim_{n \rightarrow \infty} B(\hat{\theta}) = 0$ alors l'estimateur est **consistant**. Cependant, l'inverse n'est pas vrai; un estimateur qui est **consistant** n'implique pas que la variance et le biais tendent vers 0.

Malgré la nature plaisante de la convergence d'un estimateur, beaucoup d'estimateurs ont cette propriété. Nous voulons alors une mesure qui n'indique pas seulement qu'un estimateur arrive près de la bonne valeur souvent (*alias, une très petite variance*), mais qu'il est mieux que d'autres estimateurs. De plus, dû à la sélection

arbitraire de l'erreur ϵ pour la *consistency* d'un estimateur, il est possible de malicieusement la sélectionner pour faire parler les données comme on le souhaite. Nous définissons alors l'**Erreur Quadratique Moyenne** (EQM), ou **Mean Squared Error (MSE)**, permettant de comparer les différents estimateurs ayant tous une bonne *consistency* en assurant une cohérence d'interprétation.

Erreur Quadratique Moyenne (Mean Squared Error)

$$\text{MSE}_{\hat{\theta}}(\theta) = E[(\hat{\theta} - \theta)^2 | \theta] \Leftrightarrow \text{Var}(\hat{\theta}) + [B(\hat{\theta})]^2$$

En combinant tout ces critères, le meilleur estimateur est alors l'estimateur **sans biais** ayant la **plus petite variance** possible parmi tous les estimateurs *sans biais* possible; c'est-à-dire, le **Uniformly Minimum Variance Unbiased Estimator (UMVUE)**.

Estimation par intervalles

Un type d'estimateur par intervalle est l'**intervalle de confiance** :

Intervalle de confiance

Soit le paramètre à estimer θ , alors nous sommes confiant à un niveau de $100(1 - \alpha)\%$ qu'il est contenu entre (L, U) .

De façon équivalente, nous sommes confiant à un seuil de $\alpha\%$ qu'il est contenu entre (L, U) :

$$\theta \in [L, U].$$

Nous pouvons alors dire que $\Pr(L \leq \theta \leq U) \geq (1 - \alpha)$ pour tout θ .

Par exemple, dans le cas d'une population avec distribution normale et moyenne μ inconnue, on a la moyenne échantillonnale \bar{x} (qui est l'estimateur *MVUE*).

Intervalle de confiance sur la moyenne (*distribution normale*)

Nous sommes confiant à un niveau de $100(1 - \alpha)\%$ que :

$$\mu \in \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

Construction d'estimateurs

Dans la section précédente on évalue les méthodes pour évaluer la **qualité** de l'estimateur. Cependant, comment qu'on obtient des estimateurs pour les évaluer? Plusieurs méthodes existent pour établir des estimateurs, de plus plusieurs méthodes existent pour estimer des paramètres. La méthode vu dans le cadre du

cours de statistique est la **méthode fréquentiste**, le cours de mathématiques IARD 1 (ACT-2005) présente l'**estimation Bayésienne**.

Avant de le faire nous présentons quelques concepts :

échantillon aléatoire : Échantillon d'observations indépendante provenant de la même distribution paramétrique (identiquement distribué); c'est-à-dire, un échantillon (**iid**).

k-ème moment centré à 0 : $\mu'_k = E[X^k]$.

100g^{ème} pourcentile : $\pi_g(\theta) = F_{\theta}^{-1}(g)$.

Les deux premiers estimateurs ci-dessous sont les plus faciles à obtenir, mais sont aussi les moins performants puisqu'ils n'utilisent que quelques traits des données au lieu de l'entièreté des données comme la troisième méthode.

Cette distinction devient particulièrement importante dans le cas d'une distribution avec une queue lourde à la droite (Pareto, Weibull, etc.) où il devient davantage essentiel de connaître les valeurs extrêmes pour bien estimer le paramètre de forme (α pour une Pareto).

Un autre désavantage est que les deux premières méthodes nécessitent que les données proviennent toutes de la même distribution, autrement les moments et quantiles ne seraient pas clairs.

Finalement, sous les deux premières méthodes la décision de quels moments et pourcentiles à utiliser est arbitraire.

Méthode des moments (MoM)

Soit un échantillon aléatoire de taille n (iid), on pose $\hat{\mu}'_k = \mu'_k$.

Estimation de θ par la méthode des moments

L'estimation de θ est alors toute solution des p équations :

$$\mu'_k(\theta) = \hat{\mu}'_k, \quad k = 1, 2, \dots, p$$

La raison pour cet estimateur est que la distribution empirique aura les même p premiers moments centrés à 0 que la distribution paramétrique.

Méthode du «Percentile Matching»

Soit un échantillon aléatoire de taille n (iid), on pose $\hat{\pi}_g(\theta) = \pi_g(\theta)$.

Estimation de θ par la méthode du «Percentile Matching»

L'estimation de θ est alors toute solution des p équations :

$$F(\hat{\pi}_{g_k} | \theta) = g_k, \quad k = 1, 2, \dots, p$$

La raison pour cet estimateur est que le modèle produit aura p pourcentiles qui vont «matcher» les données.

Il peut arriver que les pourcentiles de distributions ne soient pas unique, par exemple dans le cas de données discrètes lorsque le quantile recherché peut tomber entre 2 marches de la fonction empirique, ou mal-définis. Il est alors utile de définir une méthode d'interpolation des quantiles (bien qu'il n'en n'existe pas une officielle définitive).

Soit le «smoothed empirical estimate» d'un pourcentile :

Smoothed empirical estimate

On utilise les statistiques d'ordre de l'échantillon $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ pour l'interpolation suivant :

$$\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}, \quad \text{où} \\ j = \lfloor (n+1)g \rfloor \quad \text{et} \quad h = (n+1)g - j$$

Méthode du maximum de vraisemblance

Nous cherchons à maximiser la probabilité d'observer les données. Ceci est fait par la vraisemblance $\mathcal{L}(\theta; x)$ ou, puisque le logarithme ne change pas le maximum, la log-vraisemblance $\ell(\theta; x)$ où :

Maximum de vraisemblance

$$\mathcal{L}(\theta; x) = \prod_{i=1}^n f(x_i; \theta) \quad \text{et} \quad \ell(\theta; x) = \sum_{i=1}^n \ln f(x_i; \theta)$$

et l'estimateur du maximum de vraisemblance de θ est celui qui maximise la fonction de vraisemblance.

Statistiques d'ordre

Soit un échantillon aléatoire de taille n . Nous définissons la k^{e} statistique d'ordre $X_{(k)}$ comme étant la k^{e} plus petite valeur d'un échantillon. Les crochets sont utilisés pour distinguer la k^{e} statistique d'ordre $X_{(k)}$ de la k^{e} observation X_k .

Nous sommes habituellement intéressés au minimum $X_{(1)}$ et le maximum $X_{(n)}$.

Minimum

$$X_{(1)} = \min(X_1, \dots, X_n) \\ f_{X_{(1)}}(x) = n f_X(x) (S_X(x))^{n-1} \\ S_{X_{(1)}}(x) = \prod_{i=1}^n \Pr(X_i > x)$$

Maximum

$$X_{(n)} = \max(X_1, \dots, X_n) \\ f_{X_{(n)}}(x) = n f_X(x) (F_X(x))^{n-1} \\ F_{X_{(n)}}(x) = \prod_{i=1}^n \Pr(X_i \leq x)$$

De façon plus général, on défini :

k^{e} statistique d'ordre

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} \underbrace{[F_X(x)]^{k-1}}_{\text{observations} < k} \underbrace{f_X(x)}_{\text{observation} = k} \underbrace{[S_X(x)]^{n-k}}_{\text{observations} > k} \\ F_{X_{(k)}}(x) = \sum_{i=1}^n \underbrace{\binom{n}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i}}_{\text{Probabilité qu'au moins } k \text{ des } n \text{ observations } X_k \text{ sont } \leq x}$$

Nous pouvons également définir quelques autres statistiques d'intérêt :

$R = X_{(n)} - X_{(1)}$: Amplitude échantillonnale, ou **l'étendu** (Range), est la différence entre le minimum et le maximum d'un échantillon.

$M = \frac{X_{(n)} + X_{(1)}}{2}$: **mi-étendu** (Midrange), est la moyenne entre le minimum et le maximum d'un échantillon.

De plus, nous pouvons définir la **médiane** en terme de statistiques d'ordre :

$$\text{Med} = \begin{cases} X_{((n+1)/2)}, & \text{si } n \text{ est impair} \\ \frac{X_{(n/2)} + X_{(n/2+1)}}{2}, & \text{si } n \text{ est pair} \end{cases}$$

Finalement, on défini la distribution conjointe du minimum et du maximum $\forall x < y$:

$$f_{X_{(1)}, X_{(n)}}(x, y) = n(n-1)[F_X(y) - F_X(x)]^{n-2} f_X(x) f_X(y)$$