FORMULARY – RISK MODELS

Review of mathematical statistics

$$f_{Y_r}(y)=\frac{n!}{(r-1)!\,(n-r)!}\big(F_X(y)\big)^{r-1}\big(1-F_X(y)\big)^{n-r} f_X(y)$$

$$(M-m)\left(2\,f(m)\sqrt{n}\right)\overset{\circ}{\sim} n(0;1)$$

Normal populations: $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$; $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$; $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ where

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

Large samples: $Z=\dfrac{\bar{X}-\mu}{\sqrt{\operatorname{var}(\bar{X})}}\overset{\circ}{\sim}N(0,1)$; $Z=\dfrac{\bar{X}-\mu}{\sqrt{\widehat{\operatorname{var}}(\bar{X})}}\overset{\circ}{\sim}N(0,1)$

Complete data

$$S_n(x)=r_j/n,\quad y_{j-1}\le x<y_j,\quad j=2,3,\cdots,k$$

$$S_n(x)=1-\frac{c_j\,F_n(c_{j-1})-c_{j-1}\,F_n(c_j)}{c_j-c_{j-1}}-\frac{F_n(c_j)-F_n(c_{j-1})}{c_j-c_{j-1}}x,\ \ c_{j-1}\le x<c_j$$

$$S(x)=e^{-H(x)}$$

$$\hat{H}(x)=\sum_{i=1}^{j-1}(s_i/r_i),\quad y_{j-1}\le x<y_j,\quad j=2,3,\cdots,k$$

Modified Data

$$S_n(x)=\prod_{i=1}^{j-1}\left(\frac{r_i-s_i}{r_i}\right),\quad y_{j-1}\le x<y_j,\quad j=2,3,\cdots,k$$

Greenwood's formula: $\widehat{\operatorname{var}}\left(S_n(x)\right)\approx S_n(x)^2\times\sum_{i:y_i\le x}\dfrac{s_i}{r_i\left(r_i-s_i\right)}$

$$U=\exp\left(z_{\alpha/2}\,\frac{\sqrt{\widehat{\operatorname{var}}\left(S_n(x)\right)}}{S_n(x)\times\ln S_n(x)}\right)\qquad\left(\left(S_n(x)\right)^{1/U};\left(S_n(x)\right)^{U}\right)$$

$$\hat{H}(x)=\sum_{i=1}^{j-1}(s_i/r_i),\quad y_{j-1}\le x<y_j,\quad j=2,3,\cdots,k$$

$$\widehat{\operatorname{var}}\left(\hat{H}(x)\right)=\sum_{i:y_i\le x}(s_i/r_i^2)$$

$$U=\exp\left(z_{\alpha/2}\,\frac{\sqrt{\widehat{\operatorname{var}}\left(\hat{H}(x)\right)}}{\hat{H}(x)}\right)\qquad\left(U^{-1}\,\hat{H}(x);U\,\hat{H}(x)\right)$$

Kernel estimation: $\hat{f}(x)=\sum_{j=1}^{k} p(y_j)\,k_{y_j}(x)$

Maximum likelihood

(Mild regularity conditions)

$\hat{\theta}$ maximum likelihood estimator of θ. $\hat{\theta}\overset{\circ}{\sim}\text{normal}$ with mean θ and variance $I(\theta)^{-1}$. $I(\theta)\approx I(\hat{\theta})\approx -H(\hat{\theta})$;

One parameter: $I(\theta)=-n\,E\left(\frac{\partial^2}{\partial\theta^2}\ln f(X\mid\theta)\right)=-E\left(\ell''(\theta\mid X_1,X_2,\cdots,X_n)\right)$;

k parameters: $I(\theta)_{r,s}=-E\left(\frac{\partial^2}{\partial\theta_r\,\partial\theta_s}\ell(\theta\mid X_1,X_2,\cdots,X_n)\right)$ $\qquad r,s=1,2,\cdots,k$

Joint confidence interval: $c=\ell(\hat{\theta})-0.5\times q_\alpha$ and q_α quantile of a $\chi^2_{(r)}$

Bayesian estimation

$f_{Y\mid\mathbf{X}}(y\mid\mathbf{x})=\int f_{Y\mid\Theta}(y\mid\theta)\,\pi_{\Theta\mid\mathbf{X}}(\theta\mid \mathrm{x})\,d\theta$ or $f_{Y\mid\mathbf{X}}(y\mid\mathbf{x})=\sum_{\theta} f_{Y\mid\Theta}(y\mid\theta)\,\pi_{\Theta\mid\mathbf{X}}(\theta\mid \mathrm{x})$

Under regulatory conditions, $\theta\mid \mathrm{x}\overset{\circ}{\sim}\text{normal}$

Model selection

Kolmogorov-Smirnov

α	0.10	0.05	0.01
Aprox. crit. value	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

Anderson-Darling

$$A^2=-n\,F^*(u)+n\sum_{j=0}^{k}\left(1-F_n(y_j)\right)^2\left(\ln(1-F^*(y_j))-\ln(1-F^*(y_{j+1}))\right)+$$
$$+n\sum_{j=1}^{k}\left(F_n(y_j)\right)^2\left(\ln F^*(y_{j+1})-\ln F^*(y_j)\right)$$

No ties and no censoring: $A^2=-n-\sum_{j=1}^{n}\frac{2j-1}{n}\left(\ln F^*(y_j)+\ln(1-F^*(y_{n+1-j}))\right)$

α	0.10	0.05	0.01
Aprox. crit. value	1.933	2.492	3.857

Chi-squared: $Q=\sum_{j=1}^{k}\frac{\left(O_j-E_j\right)^2}{E_j}$ $\qquad Q\overset{\circ}{\sim}\chi^2_{(k-1)}$

Likelihood ratio: $\lambda(x_1,x_2,\cdots,x_n)=\frac{\sup\limits_{\theta\in\Theta_0} L(\theta\mid x_1,x_2,\cdots,x_n)}{\sup\limits_{\theta\in\Theta} L(\theta\mid x_1,x_2,\cdots,x_n)}$ and $-2\ln\lambda(X_1,X_2,\cdots,X_n)\overset{\circ}{\sim}\chi^2_{(r)}$

Simulation

Box-Muller formulae: (U_1,U_2) independent uniform variables, then $Z_1=\sqrt{-2\ln U_1}\cos(2\pi U_2)$ and $Z_2=\sqrt{-2\ln U_1}\sin(2\pi U_2)$ are independent $n(0;1)$ random variables.