Review of mathematical statistics

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

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

Normal populations:

$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)\,;\ T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}\,;$$

$$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)\ \text{where}\ S^2=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{n-1}\,;$$

$$Z=\sqrt{n-3}\left(\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right)-\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\overset{a}{\sim} n(0;1)\,;\quad \text{when } \rho=0\ \rightarrow\ T=\frac{R\sqrt{n-2}}{\sqrt{1-R^2}}\sim t(n-2)\,;$$

σ_X^2, σ_Y^2 known $\rightarrow Z=\dfrac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sqrt{\dfrac{\sigma_X^2}{n_1}+\dfrac{\sigma_Y^2}{n_2}}}\sim N(0,1)$

σ_X^2, σ_Y^2 unknown but $\sigma_X^2=\sigma_Y^2$ $\rightarrow T=\dfrac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sqrt{\dfrac{(n_1-1)\,S_X^2+(n_2-1)\,S_Y^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$

σ_X^2, σ_Y^2 unknown (Welsh approximation) $\rightarrow T=\dfrac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{S_X^2}{n_1}+\dfrac{S_Y^2}{n_2}}}\sim t(r)$, r is the largest integer

contained in $r^*=\dfrac{\left(\dfrac{s_X^2}{n_1}+\dfrac{s_Y^2}{n_2}\right)^2}{\dfrac{1}{n_1-1}\left(\dfrac{s_X^2}{n_1}\right)^2+\dfrac{1}{n_2-1}\left(\dfrac{s_Y^2}{n_2}\right)^2}$; $F=\dfrac{S_X^2}{S_Y^2}\dfrac{\sigma_Y^2}{\sigma_X^2}\sim F(n_1-1,n_2-1)$

Large samples: $Z=\dfrac{\bar{X}-\mu}{\sqrt{\operatorname{var}(\bar{X})}}\overset{\circ}{\sim} N(0,1)\,;\ Z=\dfrac{\bar{X}-\mu}{\sqrt{\widehat{\operatorname{var}}(\bar{X})}}\overset{\circ}{\sim} N(0,1)\,;\ Z=\dfrac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_X-\mu_Y\right)}{\sqrt{\dfrac{\hat{\sigma}_X^2}{n_1}+\dfrac{\hat{\sigma}_Y^2}{n_2}}}\sim N(0,1)$

Non- parametric estimation

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

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

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

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

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

Bayesian estimation

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

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: $A^2 = -n - \sum_{j=1}^{n} \dfrac{2j-1}{n} \left(\ln F^*(y_j) + \ln(1 - F^*(y_{n+1-j}))\right)$

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

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

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

Simulation

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