

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística

Relatório

Hugo Calegari RA:155738
Leonardo Uchoa Pedreira RA:156231

Professor: Verônica

Campinas-SP, 29 de Junho de 2017

Introdução Bootstrap

O método de bootstrap faz parte de uma classe de métodos não-paramétricos de Monte Carlo que estimam a distribuição de uma população ou uma característica (parâmetro de interesse) por meio de reamostragem.

Métodos de reamostragem consideram as amostras como uma população finita, a partir da qual reamostras são tomadas para estimar características e realizar inferências a respeito desta população.

Inferências baseadas em percentis pelo método de bootstrap

Ao se comparar dois grupos independentes, o método é aplicado como segue. Gera-se amostra por bootstrap para cada grupo:

- Para o j -ésimo grupo, obter amostras de bootstrap via amostragem aleatória com reposição (n_j) da seguinte amostra: X_{1j}, \dots, X_{n_j} , para obter a seguinte: $X_{1j}^*, \dots, X_{n_j}^*$;

Seja $\hat{\theta}_j^*$ a estimativa por bootstrap de θ_j , tal que este parâmetro está associado com alguma característica de interesse. Seja, ainda, $D^* = \hat{\theta}_1^* - \hat{\theta}_2^*$. Ao se repetir este processo B vezes (quantidade de réplicas) gera-se D_1^*, \dots, D_B^* . Defina $l = \frac{\alpha}{2}B$ (determinação do limite inferior do intervalo de confiança), arredonde para o inteiro mais próximo, e $u = B - l$ (limite superior). Com isso, um intervalo de confiança aproximado de $1 - \alpha$ para a diferença entre os verdadeiros parâmetros ($\theta_1 - \theta_2$) é: $[D_{(l+1)}^*, D_{(u)}^*]$, em que $D_{(1)}^* \leq \dots \leq D_{(B)}^*$.

Uma vez que se quer testar a hipótese: $H_0 : \theta_1 = \theta_2$, pode-se utilizar as seguintes estruturas de acordo com o que segue. Para as estimativas de bootstrap de $\hat{\theta}_1^*$ e $\hat{\theta}_2^*$, seja $p^* = P(\hat{\theta}_1^* > \hat{\theta}_2^*)$ (pode-se estimar esta probabilidade com o uso da proporção de $\hat{\theta}_1^* > \hat{\theta}_2^*$).

Sob a hipótese nula (igualdade dos verdadeiros parâmetros), assintoticamente (para n e B suficientemente grandes), p^* tem distribuição uniforme. Assim, rejeita-se H_0 se $p^* \leq \alpha/2$ ou se $p^* \geq 1 - \alpha/2$. Neste caso, a forma como foi estimado o valor de p^* é:

- Seja A número de valores que são maiores que zero para todos os valores das diferenças obtidos via bootstrap, isto é, entre os valores D_1^*, \dots, D_B^* . Consequentemente, pode-se estabelecer: $p^* = A/B$.

Por conveniência é adotado a o seguinte valor de p estimado: $p_m^* = \min(p^*, 1 - p^*)$ (chamado de p -valor generalizado). Com isso, rejeita-se H_0 se $p_m^* \leq \alpha/2$.

Comparação de M-estimadores

Os M-estimadores que serão avaliados são os de locação. Quando se compara estes estimadores com dois grupos independentes, ainda se percebe que a inferência baseada nos percentis por meio do método de bootstrap é o melhor método. Um intervalo de confiança baseado na estimativa do erro padrão fornecerá boa probabilidade de cobertura quando o tamanho amostral é suficientemente grande, ou seja, para se ter razoável aproximação do erro padrão necessita-se de uma população para reamostragem (amostra) relativamente grande, para que características da variabilidade populacional seja captada. A boa cobertura também depende da suposição de que as diferenças etimadas são normalmente distribuídas, porém, é desconhecido o quão grande é o tamanho amostral deveria ser antes de que a aproximação seja considerada, particularmente quando a distribuição é assimétrica.

Quando os tamanhos amostrais são pequenos, todas as indicações são de que o método de percentil por bootstrap é o melhor, então este é recomendado, até existir boa evidência de que algum outro método possa ser utilizado em seu lugar.