Estatística não paramétrica
Métodos de reamostragem

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

10 de abril de 2024

Geralmente, é relativamente simples elaborar uma estatística que mede a propriedade de interesse, mas é quase sempre difícil ou impossível determinar a distribuição dessa estatística.

Os métodos estatísticos clássicos concentraram-se principalmente nas propriedades estatísticas dos estimadores que possuem uma forma fechada simples e que podem ser analisados matematicamente. Com exceção de algumas estatísticas importantes, mas simples, esses métodos envolvem, muitas vezes, hipóteses de modelos irrealistas. Essas limitações foram superadas nas últimas duas décadas do século XX com avanços nos computadores. Uma classe de procedimentos computacionalmente intensivos conhecidos como métodos de reamostragem fornecem inferências em uma ampla gama de estatísticas sob condições muito gerais.

Métodos de reamostragem envolvem a construção de "populações" hipotéticas derivadas das observações, cada uma das quais pode ser analisada da mesma maneira para ver como as estatísticas dependem de variações aleatórias plausíveis nas observações. Reamostragem dos dados originais preserva quaisquer distribuições que estejam realmente presentes, incluindo efeitos de seleção como truncamento e censura.

Talvez o método de reamostragem mais antigo, em que se escolhe aleatoriamente repetidamente a metade dos dados e estima-se a estatística para cada amostra. A inferência sobre o parâmetro pode ser baseada no histograma da estatística reamostrada. Foi usado por Mahalanobis em 1946 sob o nome de amostras interpenetrantes. Uma variante importante é o método de Quenouille - Tukey chamado de jackknife.

Para um conjunto de n dados construímos exatamente n conjuntos de dados hipotéticos, cada um com $n - 1$ pontos, cada um omitindo um ponto diferente.

O mais importante dos métodos de reamostragem é chamado de bootstrap. Bradley Efron introduziu o método de bootstrap, também conhecido como reamostragem com substituição, em 1979. Aqui, gera-se um grande número de conjuntos de dados, cada um com n pontos de dados extraídos aleatoriamente dos dados originais. A restrição é que cada desenho é feito a partir de todo o conjunto de dados, portanto, um conjunto de dados simulado provavelmente perderá alguns pontos e terá duplicatas ou triplicatas de outros. Assim, o bootstrap pode ser visto como um método de Monte Carlo para simular dados existentes, sem qualquer suposição sobre a população.

O jackknife fornece uma abordagem de propósito geral para estimar o viés e a variância ou erro padrão de um estimador.

Suponha que $\widehat{\theta}$ seja um estimador de θ baseado na amostra aleatória $X_1, X_2, \cdots, X_n$; θ poderia ser um parâmetro desconhecido de algum modelo paramétrico ou θ poderia ser um parâmetro funcional da função de distribuição comum F dos X_i, caso em que $\theta = \theta(F)$.

O jackknife é particularmente útil quando os métodos padrão para calcular viés e variação não podem ser aplicados ou são difíceis de aplicar. Vejamos um exemplo.

Exemplo

Suponha que $X_1, \cdots, X_n$ sejam variáveis aleatórias independentes identicamente distribuídas com densidade $f(x-\theta)$ que é simétrica em torno de θ, quer dizer que $f(x) = f(-x)$. Um possível estimador de θ é a média aparada

$$\widehat{\theta} = \frac{1}{n-2g} \sum_{i=g+1}^{n-g} X_{(i)},$$

que calcula a média de $X_{(g+1)}, \cdots, X_{(n-g)}$, o meio de $n-2g$ estatísticas de ordem.

A média aparada é menos suscetível a valores extremos do que a média amostral dos X_i e é frequentemente um estimador útil de θ. No entanto, a menos que a função de densidade f seja conhecida com precisão, é difícil aproximar a variância de $\widehat{\theta}$.

Exemplo

No R, pode-se calcular a média aparada usando a função `mean()` junto com a especificação da porcentagem de corte desejada usando o parâmetro `trim`.
Veja como você pode fazer isso:

```
> set.seed(23984)
> x <- rnorm(100)
> mean(x)
[1] 0.008399732
> mean(x, trim = 0.1)
[1] 0.01635057
```

Neste exemplo, `trim = 0.1` significa que 10% dos dados serão cortados das extremidades inferior e superior antes de calcular a média. Ajuste o valor do corte conforme sua necessidade.

O nome "jackknife" foi originalmente usado por Tukey (1958) para sugerir uma técnica de ampla utilidade como um substituto para técnicas mais especializadas, da mesma forma que jackknife pode ser usado como um substituto para uma variedade de ferramentas mais especializadas embora, na realidade, um jackknife não é uma ferramenta particularmente versátil.

Referências mais completas sobre o jackknife são as monografias de Efron (1982) e Efron e Tibshirani (1993) e

Efron, Bradley and Trevor Hastie (2016). Computer Age Statistical Inference: Algorithms, Evidence, and Data Science (Institute of Mathematical Statistics Monographs). Cambridge University Press.

O estimador jackknife do vício

O estimador jackknife de viés foi desenvolvido por Quenouille (1949), embora ele não se referisse a ele como o jackknife. A idéia básica por trás dos estimadores jackknife do vício e da variância está em recomputar o estimador do parâmetro usando todas as observações.

Suponha que $\widehat{\theta}$ seja um estimador de θ com base na amostra aleatória $X_1, \cdots, X_n$, $\widehat{\theta} = \widehat{\theta}(X)$. Por exemplo, $\widehat{\theta} = \theta(\widehat{F}_n)$. O método proposto por Quenouille para estimar o vício de $\widehat{\theta}$ é baseado na exclusão sequencial de uma única observação X_i e recalculando $\widehat{\theta}$ com base em $n-1$ observações.

Suponha que

$$\mathrm{E}(\widehat{\theta}) = \theta + b(\widehat{\theta}),$$

onde $b(\widehat{\theta})$ é o vício de $\widehat{\theta}$. Seja $\widehat{\theta}_{-i}$ o estimador de θ avaliado na amostra depois de deletada X_i, ou seja,

$$\widehat{\theta}_{-i} = \widehat{\theta}(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n).$$

Definamos agora por $\widehat{\theta}_*$ a média de $\widehat{\theta}_{-1}, \widehat{\theta}_{-2}, \cdots, \widehat{\theta}_{-n}$,

$$\widehat{\theta}_* = \frac{1}{n}\sum_{i=1}^{n} \widehat{\theta}_{-i}.$$

O estimador jackknife do vício é então

$$\widehat{b}(\widehat{\theta}) = (n-1)(\widehat{\theta}_* - \widehat{\theta}).$$

A versão corrigida de viés constrói-se subtraindo-se $\widehat{b}(\widehat{\theta})$ de $\widehat{\theta}$.

Vamos mostrar abaixo que este procedimento reduz o viés de $\widehat{\theta}$.

O raciocínio teórico por trás de $\widehat{b}(\widehat{\theta})$ assume que $\mathrm{E}(\widehat{\theta})$ pode ser expressa como uma série envolvendo potências de $1/n$; por simplicidade, vamos primeiro assumir que para qualquer n

$$\mathrm{E}(\widehat{\theta}) = \theta + \frac{\alpha_1(\theta)}{n},$$

onde $\alpha_1(\theta)$ pode depender de θ ou da distribuição de X_i mas não do tamanho da amostra n; neste caso, $b(\widehat{\theta}) = \alpha_1(\theta)/n$. Dado que $\widehat{\theta}_{-i}$ é baseado em $n-1$ observações para cada i, segue que

$$\mathrm{E}(\widehat{\theta}_*) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(\widehat{\theta}_{-1}) = \theta + \frac{\alpha_1(\theta)}{n-1}.$$

Então

$$\mathrm{E}(\widehat{\theta} - \widehat{\theta}_*) = \frac{\alpha_1(\theta)}{n} - \frac{\alpha_1(\theta)}{n} = \frac{\alpha_1(\theta)}{n(n-1)},$$

e assim $n(n-1)(\widehat{\theta} - \widehat{\theta}_*)$ é um estimador não viciado de $b(\widehat{\theta})$. No caso geral, teremos

$$\mathrm{E}(\widehat{\theta}) = \theta + \frac{\alpha_1(\theta)}{n} + \frac{\alpha_2(\theta)}{n^2} + \frac{\alpha_3(\theta)}{n^3} + \cdots$$

ou

$$b(\widehat{\theta}) = \frac{\alpha_1(\theta)}{n} + \frac{\alpha_2(\theta)}{n^2} + \frac{\alpha_3(\theta)}{n^3} + \cdots,$$

onde $\alpha_1(\theta), \alpha_2(\theta), \alpha_3(\theta), \cdots$ podem depender de θ ou da distribuição de X_i, mas não de n.

Mais uma vez, segue-se que

$$\mathrm{E}(\widehat{\theta}_*) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\widehat{\theta}_{-i}) = \theta + \frac{\alpha_1(\theta)}{n-1} + \frac{\alpha_2(\theta)}{(n-1)^2} + \frac{\alpha_3(\theta)}{(n-1)^3} + \cdots,$$

lembrando que cada $\widehat{\theta}_{-i}$ depende de $n-1$ observações.

Assim, a esperança do viés do estimador jackknife é

$$\begin{aligned}\mathrm{E}\big(\widehat{b}(\widehat{\theta})\big) &= (n-1)\big(\mathrm{E}(\widehat{\theta}_*) - \mathrm{E}(\widehat{\theta})\big) \\ &= \frac{\alpha_1(\theta)}{n} + \frac{(2n-1)\alpha_2(\theta)}{n^2(n-1)} + \frac{(3n^2-3n+1)\alpha_3(\theta)}{n^3(n-1)^2} + \cdots\end{aligned}$$

Podemos ver acima que $\widehat{b}(\widehat{\theta})$ não é um estimador não-viciado de $b(\widehat{\theta})$. Assim, se definirmos

$$\widehat{\theta}_{jack} = \widehat{\theta} - b(\widehat{\theta}) = n\widehat{\theta} - (n-1)\widehat{\theta}_*$$

como sendo o estimador jackknife com correção de vício, segue que

$$\mathrm{E}(\widehat{\theta}_{jack}) \approx \theta + \frac{\alpha_2(\theta)}{n^2} - \frac{2\alpha_3(\theta)}{n^3} + \cdots$$

para n grande.

Desde que $1/n^2, 1/n^3, \cdots$ convergem a zero mais rápido do que $1/n$ tende a zero, quando n fica grande, segue-se que o viés de $\widehat{\theta}_{jack}$ é menor que o viés de $\widehat{\theta}$ para n suficientemente grande. No caso em que

$$\mathrm{E}(\widehat{\theta}) = \theta + \frac{\alpha_1(\theta)}{n},$$

de modo a $\alpha_2(\theta) = \alpha_3(\theta) = \cdots = 0$, $\widehat{\theta}_{jack}$ será não-viciado.

Exemplo

Suponhamos que $X_1, X_2, \cdots, X_n$ seja uma amostra aleatória de uma distribuição com esperança μ e variância σ^2, ambos parâmetros desconhecidos. O estimador

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

é viciado com $b(\widehat{\sigma}^2) = -\sigma^2/n$.

Para encontrar o estimador não-viciado usando o jackknife, primeiro notamos que

$$\overline{X}_{-i} = \frac{1}{n-1}\sum_{j\neq i} X_j = \frac{1}{n-1}(n\overline{X} - X_i).$$

Então

$$\begin{aligned}\widehat{\sigma}^2_{-i} &= \frac{1}{n-1}\sum_{j\neq i}(x_j - \overline{x}_{-i})^2 \\ &= \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2 - \frac{n}{(n-1)^2}(x_i - \overline{x})^2.\end{aligned}$$

Agora, $\widehat{\sigma}^2_*$ é justamente a média dos $\widehat{\sigma}^2_{-i}$, de maneira que

$$\widehat{\sigma}^2_* = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 - \frac{1}{(n-1)^2}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

e o estimador não viciado de σ^2 é

$$n\widehat{\sigma}^2 - (n-1)\widehat{\sigma}^2_* = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = s^2.$$

Exemplo

Suponhamos que $X_1, X_2, \cdot, X_n$ seja uma amostra aleatória com função de densidade

$$f(x;\theta) = \frac{1}{\theta}, \quad \text{para} \quad 0 \leq x \leq \theta,$$

onde θ é um parâmetro desconhecido. Desde que θ é o valor máximo possível dos X_i, o estimador natural de θ é

$$\widehat{\theta} = X_{(n)} = \max(X_1, X_2, \cdots, X_n).$$

No entanto, como os X_i não podem exceder θ, segue que o máximo deles não pode exceder θ e, portanto, $\widehat{\theta}$ é viciado; de fato,

$$\mathrm{E}(\widehat{\theta}) = \frac{n}{n+1}\theta = \theta\frac{1}{1+1/n} = \theta\left(1 - \frac{1}{n} + \frac{1}{n^2} - \frac{1}{n^3} + \cdots\right).$$

Dado que

$$\widehat{\theta}_{-i} = \max(X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n),$$

observamos que $\widehat{\theta}_{-i} = X_{(n)}$ para $n-1$ valores de i e $\widehat{\theta}_{-i} = X_{(n-1)}$ para os outros valores de i. Então, obtemos

$$\widehat{\theta}_* = \frac{n-1}{n} X_{(n)} + \frac{1}{n} X_{(n-1)}$$

e o estimador jackknife de θ seria

$$\widehat{\theta}_{jack} = X_{(n)} + \frac{n-1}{n}\left(X_{(n)} - X_{(n-1)}\right).$$

O vício de $\widehat{\theta}_{jack}$ será menor que o de $\widehat{\theta}$; no entanto, podemos facilmente modificar $\widehat{\theta}$ para torná-lo não-viciado sem recorrer ao jackknife, simplesmente multiplicando-o por $(n+1)/n$.

Este ultimo exemplo aponta uma das desvantagens de usar qualquer método de propósito geral como o jackknife, ou seja, em situações específicas muitas vezes é possível melhorar o estimador com um método adaptado especificamente para a situação em questão.

Remover o viés em $\widehat{\theta} = X_{(n)}$ multiplicando-o por $(n+1)/n$ depende do fato de que a forma da densidade seja conhecida. Suponha, em vez disso, que o intervalo dos X_i ainda seja $[0, \theta]$, mas que a densidade f seja desconhecida para $0 \leq x \leq \theta$. Então $X_{(n)}$ ainda é um estimador razoável de θ e sempre o subestima. No entanto, $(n+1)X_{(n)}/n$ não precisa ser não viciado e, de fato, pode ser mais severamente influenciado do que $X_{(n)}$. No entanto, o estimador jackknife

$$\widehat{\theta}_{jack} = X_{(n)} + \frac{n-1}{n}\left(X_{(n)} - X_{(n-1)}\right),$$

terá um viés menor que $X_{(n)}$ e pode ser preferível.

O estimador jackknife da variância

O estimador jackknife de vício usa os estimadores $\widehat{\theta}_{-1}, \cdots, \widehat{\theta}_{-n}$, que usam todas as observações, exceto uma em seu cálculo para construir um estimador de viés de um estimador $\widehat{\theta}$. Tukey (1958) sugeriu um método para estimar Var($\widehat{\theta}$).

O estimador jackknife de Tukey de Var($\widehat{\theta}$) é

$$\widehat{\text{Var}}(\widehat{\theta}) = \frac{n-1}{n}\sum_{i=1}^{n}(\widehat{\theta}_{-i} - \widehat{\theta}_{*})^2,$$

onde, como antes $\widehat{\theta}_{-i}$ é o estimador avaliado usando todas as observações, exceto X_i e

$$\widehat{\theta}_{*} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta}_{-i}.$$

A fórmula para o estimador da variância jackknife é um tanto não intuitiva. Ao derivar a fórmula, Tukey assumiu que o estimador $\widehat{\theta}$ pode ser bem aproximado por uma média de variáveis aleatórias independentes; essa suposição é válida para uma ampla variedade de estimadores, mas não é verdadeira para alguns estimadores, por exemplo, máximo ou mínimo de amostra. Mais precisamente, Tukey assumiu que

$$\widehat{\theta} \approx \frac{1}{n}\sum_{i=1}^{n}\phi(X_i),$$

o qual sugere que

$$\mathrm{Var}\widehat{\theta} \approx \frac{1}{n}\mathrm{Var}\big(\phi(X_1)\big).$$

No caso em que o parâmetro de interesse θ é um parâmetro funcional da função de distribuição F, isto é, $\theta = \theta(F)$, a função $\phi(\cdot) - \theta(F)$ é tipicamente a curva de influência de $\theta(F)$.

Em geral, não conhecemos a função $\phi(x)$, portanto não podemos usar diretamente a fórmula acima. No entanto, é possível encontrar substitutos razoáveis para $\phi(X_1), \cdots, \phi(X_n)$. Usando os estimadores $\widehat{\theta}_{-i}$, $i = 1, \cdots, n$ e $\widehat{\theta}$, definimos pseudovalores

$$\Phi_i = \widehat{\theta} + (n-1)(\widehat{\theta} - \widehat{\theta}_{-i}),$$

para $i = 1, \cdots, n$ que essencialmente desempenham o mesmo papel que o $\phi(X_i)$ acima.

No caso em que $\theta = \theta(F)$, $(n-1)(\widehat{\theta} - \widehat{\theta}_{-i})$ é uma tentativa de estimar a curvatura de $\theta(F)$ em $x = X_i$. No caso em que $\widehat{\theta}$ é exatamente a média amostral

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i),$$

é fácil mostrar que $\Phi_i = \phi(X_i)$ e assim a conexão entre Φ_i e $\phi(X_i)$ é clara neste caso simples.

Podemos então tomar a variância da amostra dos pseudovalores Φ_i como sendo uma estimativa da variância de $\phi(X_1)$ e usá-lo para estimar a variância de $\widehat{\theta}$. Note que

$$\frac{1}{n}\sum_{i=1}^{n}\Phi_i = n\widehat{\theta} - \frac{n-1}{n}\sum_{i=1}^{n}\widehat{\theta}_{-i} = n\widehat{\theta} - (n-1)\widehat{\theta}_* = \widehat{\theta}_{jack},$$

onde $\widehat{\theta}_{jack}$ é a versão de $\widehat{\theta}$ com correção de viés.

A variância amostral dos Φ_i é

$$\frac{1}{n-1}\sum_{i=1}^{n}(\Phi_i - \phi)^2 = \cdots = (n-1)\sum_{i=1}^{n}(\widehat{\theta}_* - \widehat{\theta}_{-i})^2.$$

Agora obtemos o estimador jackknife da variância dividindo a variância amostral de cada Φ_i por n:

$$\widehat{\text{Var}}(\widehat{\theta}) = \frac{n-1}{n} \sum_{i=1}^{n} (\widehat{\theta}_* - \widehat{\theta}_{-i})^2.$$

Deve-se notar que o estimador jackknife da variância não funciona em todas as situações. Uma dessas situações é a mediana amostral. O problema aqui parece ser o fato de que a curva de influência da mediana é definida apenas para distribuições contínuas e, portanto, é difícil de aproximar adequadamente em amostras finitas.

Exemplo

Os dados da tabela representam uma amostra de 30 rendimentos antes dos impostos. Vamos supor que esses dados sejam resultados da amostra aleatória $X_1, \cdots, X_{30}$ com função de distribuição F. Vamos usar os dados para estimar o índice de Gini

$$\theta(F) = 1 - 2\int_0^1 q_F(t)\mathrm{d}t,$$

onde

$$q_F(t) = \frac{\int_0^t F^{-1}(s)\mathrm{d}s}{\int_0^1 F^{-1}(s)\mathrm{d}s},$$

é a curva de Lorenz.

3841	7084	7254	15228	18042	19089
22588	23972	25694	27592	27927	31576
32528	32921	33724	36887	37776	37992
39464	40506	44516	46538	51088	51955
54339	57935	75137	82612	83381	84741

Rendimentos antes dos impostos.

Utilizando o princípio de substituição o estimador de $\theta(F)$ é

$$\widehat{\theta} = \theta(\widehat{F}_n) = \frac{1}{\sum_{i=1}^{30} X_i} \sum_{i=1}^{30} \left(\frac{2i-1}{30} - 1 \right) X_{(i)},$$

sendo $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(30)}$ as estatística de ordem.

Para os dados na tabela, a estimativa de $\theta(\widehat{F}_n)$ é 0.311. As estimativas dos $\widehat{\theta}_{-i}$ de $\theta(\widehat{F}_n)$ são dadas na tabela abaixo.

0.2912	0.2948	0.2950	0.3028	0.3055	0.3064
0.3092	0.3103	0.3115	0.3127	0.3129	0.3148
0.3153	0.3154	0.3157	0.3166	0.3168	0.3168
0.3170	0.3170	0.3169	0.3167	0.3161	0.3159
0.3152	0.3140	0.3069	0.3033	0.3028	0.3020

Valores de $\hat{\theta}_{-i}$ obtidos por omissão da entrada correspondente na tabela anterior.

O estimador jackknife do desvio padrão de $\widehat{\theta}$ é

$$\widehat{\mathrm{se}}(\widehat{\theta}) = \frac{29}{30}\sum_{i=1}^{30}(\widehat{\theta}_* - \widehat{\theta}_{-i})^2 = 0.0398,$$

onde $\widehat{\theta}_* = 0.310$ é a média de $\widehat{\theta}_{-1}, \cdots, \widehat{\theta}_{-30}$.

A importância do bootstrap surgiu durante a década de 1980, quando o estudo matemático demonstrou que ele fornece estimativas quase ótimas da distribuição de muitas estatísticas sob uma ampla gama de circunstâncias.

Em vários casos, o método produz melhores resultados do que aqueles obtidos pela teoria clássica da aproximação normal.

No entanto, deve-se alertar que o bootstrap não é a solução para todos os problemas. A teoria desenvolvida nas décadas de 1980 e 1990 mostra que o bootstrap falha em algumas situações "não-suaves". Por isso, deve-se ter cautela e resistir à tentação de usar o método de forma inadequada.

Como os computadores se tornaram mais poderosos, o bootstrap, assim como os procedimentos de reamostragem em geral ganharam uso generalizado. O bootstrap é uma ferramenta geral usada para medir o erro em uma estimativa ou a significância de um teste de hipótese.

Aqui demonstramos a utilidade do bootstrap para uma variedade de problemas, embora ainda apenas arranhemos a superfície; o leitor interessado em um tratamento completo é referido a Efron e Tibshirani (1993) ou o de Davison e Hinkley (1997). Ilustramos a como encontrar intervalos de confiança e p-valores para os problemas de posição de uma amostra e outros.

Descrição do método bootstrap

Para fixar ideias, lembre-se que o histograma da amostra é frequentemente usado para fornecer o contexto da distribuição da variável aleatória, por exemplo, localização, variabilidade, forma. Uma maneira de pensar no bootstrap é como um procedimento para fornecer algum contexto para a distribuição amostral de uma estatística. Uma amostra bootstrap é simplesmente uma amostra da amostra original obtida com reposição.

A ideia é que, se a amostra é representativa da população, ou mais concretamente, se o histograma ou o estimador kernel da função de densidade da amostra se assemelha à função de densidade da variável aleatória, então a amostragem da amostra é representativa da amostragem da população. Fazê-lo repetidamente produzirá uma estimativa da distribuição amostral da estatística.

Seja $X = (X_1, \cdots, X_n)$ uma amostra aleatória extraída de uma distribuição populacional desconhecida F. Suponha que $T_n(X)$ seja um bom estimador de $T(F)$, um parâmetro de interesse.

O interesse reside em avaliar sua precisão na previsão. Determinar os intervalos de confiança para $T(F)$ requer conhecimento da distribuição amostral G_n de $T_n(X) - T(F)$, isto é,

$$G_n(x) = P\big(T_n(X) - T(F) \leq x\big),$$

para todo x.

Por exemplo, a média amostral $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ é um bom estimador da esperança populacional μ. Para obter o intervalo de confiança para μ, devemos encontrar a distribuição amostral de $\overline{X} - \mu$, que depende da forma e outras características da distribuição desconhecida F.

A teoria estatística clássica usa o Teorema do Limite Central como aproximação normal para a distribuição amostral.

Por exemplo, se $(X_1, Y_1), \cdots, (X_n, Y_n)$ denotam observações de uma população normal bivariada, então o estimador de máxima verossimilhança do coeficiente de correlação ρ é dado pelo coeficiente de correlação de Pearson amostral

$$\widehat{\rho} = \frac{\sum_{i=1}^{n}(X_i Y_i - \overline{XY})}{\sqrt{\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right)\left(\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right)}}.$$

Para estatísticas com distribuições assimétricas, como a de $\widehat{\rho}$, a teoria clássica sugere transformações de variáveis. Neste caso, a transformação Z de Fisher dada por

$$Z = \frac{\sqrt{n-3}}{2}\left(\ln\left(\frac{1+\widehat{\rho}}{1-\widehat{\rho}}\right) - \ln\left(\frac{1+\rho}{1-\rho}\right)\right),$$

dá uma melhor aproximação à normal.

Esta aproximação corrige a assimetria e é melhor que a aproximação à normal de $\sqrt{n}(\widehat{\rho} - \rho)$. O método de bootstrap, quando usado adequadamente, evita transformações levando em conta a assimetria da distribuição amostral.

O método de bootstrap presume que, se $\widehat{F}_n$ é uma boa aproximação da distribuição da população desconhecida F, então o comportamento das amostras de $\widehat{F}_n$ se assemelha ao dos dados originais.

Aqui $\widehat{F}_n$ pode ser a função de distribuição empírica da amostra aleatória $X_1, \cdots, X_n$ ou um estimador paramétrico da função F.

Uma vez que $\widehat{F}_n$ é fornecido, os conjuntos de dados

$$X^* = (X_1^*, \cdots, X_n^*)$$

são reamostrados a partir de $\widehat{F}_n$ e a estatística $T_n(X^*)$ é computada para cada reamostra.

Sob condições muito gerais pode-se demonstrar que a diferença entre a distribuição amostral G_n de $T_n(X) - T(F)$ e a distribuição bootstrap G_b, isto é, a distribuição de $T_n(X^*) - T(\widehat{F}_n)$ é insignificante. G_b pode ser usado para extrair inferências sobre o parâmetro $T(F)$ no lugar do desconhecido G_n.

Em princípio a distribuição de bootstrap, obtida pelo histograma ou pelo estimador kernel de densidades, G_b é completamente conhecida, pois é construída inteiramente a partir dos dados originais. No entanto, para obter a distribuição completa de bootstrap, é necessário calcular as estatísticas para quase todas as n^n amostras possíveis bootstrap.

Para o exemplo simples da média amostral, presumivelmente, é necessário computar

$$\begin{array}{ll} X_1^{*(1)}, \cdots, X_n^{*(1)}, & r_1 = \overline{X}^{*(1)} - \overline{X} \\ X_1^{*(2)}, \cdots, X_n^{*(2)}, & r_2 = \overline{X}^{*(2)} - \overline{X} \\ \qquad \ddots \quad \ddots & \ddots \\ X_1^{*(n^n)}, \cdots, X_n^{*(n^n)}, & r_{n^n} = \overline{X}^{*(n^n)} - \overline{X} \end{array}$$

A distribuição bootstrap dada pelo histograma ou pelo estimador kernel de densidade de $r_1, \cdots, r_{n^n}$. Mesmo para $n = 10$ dados, n^n acaba por ser dez bilhões. Na prática, a estatística de interesse, $T_n(X^*) - T(\widehat{F}_n)$ é calculada para um número N, digamos $N = n \log^2(n)$, de novas amostras e sua distribuição bootstrap construída.

O bootstrap mais popular e simples é o bootstrap não paramétrico, onde a reamostragem com reposição é baseada na função de distribuição empírica da amostra aleatória original. Isto dá pesos iguais para cada um dos dados originais.

A tabela abaixo fornece versões bootstrap de algumas estatísticas comumente usadas. No caso do estimador de razão e do coeficiente de correlação, os pares de dados são reamostrados a partir dos pares de dados originais (X_i, Y_i).

Estatística	Versão bootstrap
Média: $\overline{X}$	$\overline{X}^*$
Variância: $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$	$\frac{1}{n}\sum_{i=1}^{n}(X_i^* - \overline{X}^*)^2$
Estimador da razão: $\frac{\overline{X}}{\overline{Y}}$	$\frac{\overline{X}^*}{\overline{Y}^*}$
Coeficiente de correlação: $\frac{\sum_{i=1}^{n}(X_iY_i - \overline{X}\overline{Y})}{\sqrt{(\sum_{i=1}^{n}(X_i - \overline{X})^2)(\sum_{i=1}^{n}(Y_i - \overline{Y})^2)}}$	$\frac{\sum_{i=1}^{n}(X_i^*Y_i^* - \overline{X}^*\overline{Y}^*)}{\sqrt{(\sum_{i=1}^{n}(X_i^* - \overline{X}^*)^2)(\sum_{i=1}^{n}(Y_i^* - \overline{Y}^*)^2)}}$

Estatísticas e suas versões bootstrap.

Intervalos confidenciais bootstrap

O conceito de quantidade pivotal é muito importante para intervalos de confiança. Por exemplo, o fator de escala $\overline{X} - \mu$ depende do desvio padrão σ ou de seu estimador. É possível obter intervalos de confiança ou intervalos aproximados de confiança para μ, se $P\left(\overline{X} - \mu\right)/\sigma < x)$ não depende de μ e σ.

A menos que a distribuição limite esteja livre dos parâmetros desconhecidos, uma vez que não pode ser invertida para obter intervalos de confiança.

Assim, é importante concentrar-se em quantidades pivotais ou aproximadamente pivotais, a fim de obter intervalos de confiança confiáveis para o parâmetro de interesse.

Uma função $T_n(X;F)$ é pivotal, se sua distribuição for livre de parâmetros desconhecidos de F. No caso, $X_i \sim N(\mu, \sigma^2)$, então $T_n(X;F) = \sqrt{n}(\overline{X} - \mu)/s_n$ é pivotal. No caso não normal, é aproximadamente pivotal. Para obter o intervalo de confiança bootstrap para μ, calculamos $\sqrt{n}(\overline{X}^{*(j)} - \overline{X})/s_n$ para N amostras bootstrap e organizamos os valores em ordem crescente

$$h_1 \leq h_2 \leq \cdots \leq h_N.$$

Pode-se então ler do histograma (digamos) o intervalo de confiança de 90% do parâmetro. Ou seja, o intervalo de confiança de 90% para μ é dado por

$$\overline{X} - h_m \frac{s_n}{\sqrt{n}} \leq \mu \leq \overline{X} - h_k \frac{s_n}{\sqrt{n}},$$

onde $k = [0.5N]$ e $m = [0.95N]$. Babu Singh (1983) mostraram que $N = n \log^2(n)$ iterações bootstrap seriam suficientes.

Exemplo

Para ilustrar o uso do bootstrap, primeiro geramos uma amostra de tamanho 25 a partir de uma distribuição normal com esperança 30 e desvio padrão 5.

```
> set.seed(8473)
> x = rnorm(25,30,5)
```

No código a seguir obtemos 1000 amostras bootstrap e, para cada amostra, calculamos a média da amostra. O vetor resultante contém as 1000 médias da amostra. Mostramos também o histograma das 1000 estimativas. Incluímos também um gráfico da verdadeira função de densidade da distribuição amostral de $\overline{X} \sim N(30, 5^2/25)$.

Exemplo

```
> B = 1000 # número de amostras bootstrap a serem obtidas
> xbar = rep(0,B)
> for( i in 1:B ) {
   xbs = sample(x,length(x), replace=TRUE)
   xbar[i] = mean(xbs)}
```

O desvio padrão da distribuição amostral bootstrap pode servir como uma estimativa do erro padrão da estimativa.

```
> se.xbar = sd(xbar)
> se.xbar
[1] 0.9776717
```

O erro padrão estimado pode então ser usado para inferência. Como sabemos que a distribuição da média amostral é normal, podemos calcular um intervalo de confiança aproximado de 95% usando os valores t críticos como segue. Incluímos o intervalo t usual para comparação.

Exemplo

```
> tcv = qt(0.975,length(x)-1)
> mean(x)+c(-1,1)*tcv*se.xbar
[1] 30.13485 34.17048
> mean(x)+c(-1,1)*tcv*sd(x)/sqrt(length(x))
[1] 30.13744 34.16788
```

Uma outra forma é utilizarmos as funções no pacote bootstrap.

```
> library(bootstrap)
> set.seed(8473)
> x = rnorm(25,30,5)
> resultados = bootstrap(x, 1000, mean)
> mean(resultados$thetastar)
[1] 32.15428
> sd(resultados$thetastar)
[1] 0.9619809
```

Testes Bootstrap de hipóteses

No teste bootstrap, a reamostragem é conduzida sob condições que garantem que a hipótese nula H_0, seja verdadeira. Isso permite a formulação de um p-valor bootstrap.

O procedimento de teste de bootstrap para o problema emparelhado é o seguinte. Primeiro, amostra-se com reposição do conjunto de pares; então o tratamento é aplicado aleatoriamente ao par. Observe que isso preserva a correlação do plano emparelhado. Se $d_1, \cdots, d_n$ denotam as diferenças com base nos dados da amostra original, então na amostra bootstrap, se o i-ésimo par for selecionado d_i e d_i cada um tem probabilidade 1/2 de estar na amostra bootstrap; portanto, a hipótese nula é verdadeira.

Sejam $T_1^*, \cdots, T_B$ as estatísticas de teste baseadas nas B amostras bootstrap. Estes formam uma estimativa da distribuição nula da estatística de teste T. O p-valor bootstrap é então calculado como

$$p\text{-valor} = \frac{\#\{T_i^* \geq T\}}{B}.$$

Exemplo: Intervenção em Creche.

Este conjunto de dados é extraído de um estudo discutido por Siegel (1956). Envolve oito pares de gêmeos idênticos em idade escolar. No estudo, para cada par, um é selecionado aleatoriamente para frequentar a creche enquanto o outro permanece em casa.
No final do período de estudo, todas as 16 crianças passam pelo mesmo teste de consciência social. Para cada par, a resposta de interesse é a diferença nas pontuações dos gêmeos, Gêmeo na Escola - Gêmeo em Casa.

Exemplo: Intervenção em Creche.

```
> school<-c(82,69,73,43,58,56,76,65)
> home<-c(63,42,74,37,51,43,80,62)
> d<-school-home
> dpm<-c(d,-d)
```

Então, o vetor **dpm** contém todas as $2n$ diferenças possíveis. A obtenção de amostras bootstrap deste vetor garante que a hipótese nula seja verdadeira. A seguir, primeiro obtemos $B = 5000$ amostras bootstrap e as armazenamos no vetor **dbs**.

```
> n<-length(d)
> B<-5000
> dbs<-matrix(sample(dpm,n*B,replace=TRUE),ncol=n)
```

Exemplo: Intervenção em Creche.

A seguir usaremos a função **apply** para obter a estatística do teste de Wilcoxon para cada amostra bootstrap. Primeiro definimos uma função que retornará o valor da estatística de teste.

```
> wilcox.teststat<-function(x) wilcox.test(x)$statistic
> bs.teststat<-apply(dbs,1,wilcox.teststat)
> mean(bs.teststat>=wilcox.teststat(d))
[1] 0.0238
```

Portanto, o *p*-valor = 0.0238 e é significativo ao nível de 5%.