Application of Monte Carlo Methods: The Bootstrap

Statistical inference

- Suppose we have a sample $x_1, \ldots, x_n$.
- We assume that $x_1, \ldots, x_n$ is a (iid) realizations of random variables $X_1, \ldots, X_n$.
- We then assume that the distribution of the random sample depends on some unknown quantity θ_0 and we want to estimate θ_0.
- We do this by postulating a statistical model $\{f_\theta, \theta \in \Theta\}$, and assume that $\theta_0 \in \Theta$.
- From the model we propose an estimator $\hat{\theta} = T(X_1, \ldots, X_n)$. And we evaluate the estimator on the data to get an estimate $T(x_1, \ldots, x_n)$ of θ_0.

Statistical inference

Example:

- For instance we could assume that $X_1, \ldots, X_n$ are iid $N(\mu_0, 1)$ and we want to estimate the mean μ_0. (μ_0 represents θ_0 above).
- We propose the estimator

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

- Note that $\hat{\mu}$ is a random variable. Its value on the observed data is $\frac{1}{n} \sum_{i=1}^{n} x_i$, which is our estimate.
- We wish to derive some basic properties of this estimator.

Statistical inference

Definition

- The <u>bias</u> of $\hat{\theta} = T(X_1, \ldots, X_n)$ is the quantity:

$$\mathrm{B}(\hat{\theta}) = \mathbb{E}\left(\hat{\theta}\right) - \theta_0.$$

  If $\mathrm{B}(\hat{\theta}) = 0$, we say that $\hat{\theta}$ is an <u>unbiased estimator</u> of θ_0.

- The variance of $\hat{\theta}$ is

$$\mathrm{Var}\left(\hat{\theta}\right) = \mathbb{E}\left[\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right)^2\right].$$

Statistical inference

- On average, an unbiaised estimator gives the correct estimate θ_0. This mean that we get exactly θ_0, by averaging the values of the estimator over a very large number of sampling experiments (law of large numbers).
- But we typically have a single random sample. So an unbiased estimator can be substantially off if its variance is large.
- Often we are willing to accept some bias for less variance. One way to quantity this trade-off is to seek estimators with the smallest Mean Square Error (MSE) defined as

$$\mathsf{MSE}(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta} - \theta_0\right)^2\right].$$

Statistical inference

Proposition

$$MSE(\hat{\theta}) = Var\left(\hat{\theta}\right) + \left(B(\hat{\theta})\right)^2.$$

- When comparing estimators, we will compare their MSE.
- If two estimators are unbiased, this is equivalent to comparing their variances.
- The question is: how do we compute the bias and the MSE? In few cases this is easy to do. In most cases, it is complicated, and Monte Carlo methods can be helpful in this task.

Parametric bootstrap

Example:

- Suppose that $X_1, \ldots, X_n$ are iid $N(\mu_0, 1)$ and $\hat{\mu} = n^{-1} \sum_{i=1}^{n} X_i$. This estimator is well understood. $B(\hat{\mu}) = 0$ and $\text{MSE}(\hat{\mu}) = n^{-1}$. No calculation needed.

Another example:

- Suppose that $X_1, \ldots, X_n$ are iid $\text{Exp}(\lambda_0)$ ($f_\lambda(x) = \lambda e^{-\lambda x}$, $x \geq 0$) and we form $\hat{\lambda} = n / \sum_{i=1}^{n} X_i$.
- Clearly $B(\hat{\lambda}) \neq 0$. How to calculate $B(\hat{\lambda})$ and $\text{MSE}(\hat{\lambda})$? The challenge is that $B(\hat{\lambda})$ and $\text{MSE}(\hat{\lambda})$ are complicated integrals that depend on the unknown parameter θ_0.

Parametric bootstrap

- We now discuss a general method called bootstrap that can be used to estimate bias and MSE (among other things).
- We consider again the general case where $X_1, \dots, X_n$ are iid random sample from a distribution f_{θ_0} and we are interested in estimating θ_0.
- Given an estimator $\hat{\theta} = T(X_1, \dots, X_n)$, its bias and MSE are

$$B(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta_0, \quad \text{and} \quad \mathrm{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta_0)^2\right].$$

Parametric bootstrap

- There are two main difficulties in estimating $B(\hat{\theta})$: 1) θ_0 is not known, and 2) the expectations in these terms may not be tractable.
- This becomes clear by noting that, since $\hat{\theta} = T(X_1, \ldots, X_n)$, we have

$$\begin{aligned} B(\hat{\theta}) &= \mathbb{E}\left(T(X_1, \ldots, X_n)\right) - \theta_0, \\ &= \int T(x_1, \ldots, x_n) f_{\theta_0}(x_1) \cdots f_{\theta_0}(x_n) dx_1 \cdots dx_n - \theta_0. \end{aligned}$$

- Similarly for the variance.

Parametric bootstrap

- The idea of Bootstrap is to estimate θ_0 from the data and to use Monte Carlo calculation to approximate the expectation $\mathbb{E}(\hat{\theta})$ (the integral).
- Because we assumed that $X_i \sim f_{\theta_0}$ (a parametric family), the method discussed above is known as the parametric bootstrap.

Parametric bootstrap

Example:

- Suppose that $X_1, \dots, X_n$ are iid $\text{Exp}(\lambda_0)$ ($f_{\lambda_0}(x) = \lambda_0 e^{-\lambda_0 x}$, $x \geq 0$) and we are interested in estimating λ_0.
- We consider $\hat{\lambda} = n / \sum_{i=1}^n X_i$.

$$
\begin{aligned}
B(\hat{\lambda}) &= \mathbb{E}(\hat{\lambda}) - \lambda_0 \\
&= \int \frac{n}{\sum_{i=1}^n x_i} f_{\lambda_0}(x_1) \cdots f_{\lambda_0}(x_n) dx_1 \cdots dx_n - \lambda_0.
\end{aligned}
$$

- Here we clearly see the two issues: we don't know λ_0; and even if we do, we don't know how to compute the big integral.

Parametric bootstrap

Example:

- The bootstrap addresses these two issues in a very simple way.
- We estimate λ_0 in $\mathbb{E}(\hat{\lambda}) - \lambda_0$ by $\hat{\lambda}$, and we estimate the expectation $\mathbb{E}(\hat{\lambda})$ by Monte Carlo, by sampling from $f_{\hat{\lambda}_0}$.
- We proceed similarly for the MSE.

Parametric bootstrap

This gives the following algorithm.

Algorithm

Let B be the number of bootstrap replications.

1. *Estimate λ_0 from the data by $\hat{\lambda} = n / \sum_{k=1}^{n} x_k$.*
2. *For $i = 1, \ldots, B$, generate $(X_1^{(i)}, \ldots, X_n^{(i)})$ an iid sample of size n from $f_{\hat{\lambda}}$. For each sample, calculate*

$$\hat{\lambda}^{(i)} = \frac{n}{\sum_{k=1}^{n} X_k^{(i)}}.$$

3. *Estimate $B(\hat{\lambda})$ by*

$$\frac{1}{B} \sum_{i=1}^{B} \hat{\lambda}^{(i)} - \hat{\lambda}.$$

Parametric bootstrap

```
lambda=2 #true value
n=30 #sample size
dt=rexp(n,lambda) # Data set
B=200 #Number of Bootstrap replication
hatlbd=1/mean(dt);Vec=numeric(B)
for (i in 1:B){
   Vec[i]=1/mean(rexp(n,hatlbd))
}
hist(Vec-hatlbd,20)
mean(Vec-hatlbd)
#This is the bootstrap estimate of the bias
[1] 0.067
```

Parametric bootstrap

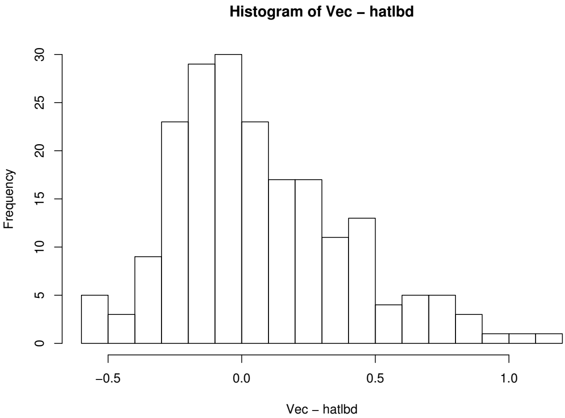

Figure 1: Histogram of $\hat{\lambda}^{(i)} - \hat{\lambda}$.

Parametric bootstrap

Practice: Write similar code to estimate the MSE.

Nonparametric bootstrap

- Looking at the above example, we can understand bootstrap as follows. Say we have a sample $x_1, \ldots, x_n$ from some random sample with unknown cumulative distribution function (cdf) F_{θ_0} and we want to estimate the bias of some estimator $\hat{\theta}$ of θ_0.
- Clearly, if we could obtain many other random samples from F_{θ_0}, we could evaluate the estimator on each of them and use the sample bias to approximate the population bias of the estimator.
- But in practice, we have only one sample. The idea of bootstrap is to use that single sample to estimate the distribution F_{θ_0} so that the idea above can carry through.

Nonparametric bootstrap

- When F_{θ_0}, the unknown distribution belong to a parametric family, estimating F_{θ_0} boils down to estimating θ_0.
- This approach is called parametric bootstrap.
- Another possibility is to estimate the population cdf F_{θ_0} by the empirical cdf and proceed similarly.
- This is called nonparametric bootstrap.

Nonparametric bootstrap

- Let $X_1, \ldots, X_n$ be iid F. The empirical cdf is defined as:

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{(-\infty, x]}(X_i) = \; = \quad \frac{\#\{1 \leq i \leq n : \; X_i \leq x\}}{n}.$$

- This estimate makes sense because $F(x) = \mathbb{P}(X \leq x)$. Hence, as $n \to \infty$, $\hat{F}(x) \to F(x)$, by the law of large numbers.
- Note that $\hat{F}$ is simply the cdf of the discrete probability distribution that put mass $1/n$ on each of the data point.
- The following R code compares a true and an empirical cdf for the standard normal distribution $N(0, 1)$.

Nonparametric bootstrap

```
n=100;X=rnorm(100) # random sample
grid=seq(-4,4,length=200) # grid where to evaluate cdf
CDF1=pnorm(grid);
for (i in 1:200){
   CDF2[i]=sum(X<=grid[i])/n
}
plot(grid,CDF2,type='s',col='blue',xlab='',
      ylab='',main='empirical versus true CDF')
par(new=T)
plot(grid,CDF1,type='l',col='red', ,xlab='',
       ylab='',main='empirical versus true CDF')
```

Nonparametric bootstrap

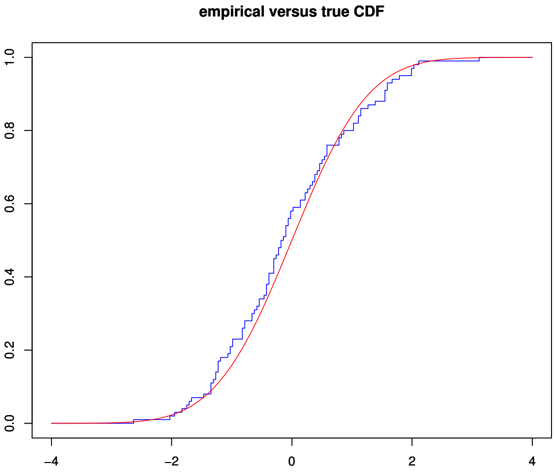

Figure 3: cdf and empirical cdf for the $N(0, 1)$ based on 100 random points.

Nonparametric bootstrap

Replacing F by $\hat{F}$, the bootstrap algorithm then becomes.

Algorithm

Set B the number of bootstrap replication

1. *For $i = 1$ to B: generate $X_1^{(i)}, \dots, X_n^{(i)}$ iid from $\hat{F}$ and using each sample, calculate $\hat{\theta}^{(i)} = T(X_1^{(i)}, \dots, X_n^{(i)})$.*
2. *Estimate the bias by $\frac{1}{B}\sum_{i=1}^{B} \hat{\theta}^{(i)} - \hat{\theta}$.*
3. *To estimate the mean square error, use $\frac{1}{B}\sum_{i=1}^{B}(\hat{\theta}^{(i)} - \hat{\theta})^2$.*

- note that we do not need to explicitly compute the estimated cdf $\hat{F}$. We only need to generate samples from it.

Nonparametric bootstrap

- There is one small detail: we need to sample from $\hat{F}$. Since $\hat{F}$ is a discrete probability distribution, we can easily use the inversion method: generate $U \sim \mathcal{U}(0,1)$ and return X_I, where I is the first k s.t. $\sum_{i=1}^{k} i/n \geq U$.
- But more simply, we can sample from $\hat{F}$ using:

  Generate $U \sim \mathcal{U}(1, n+1)$. Set $I = \lfloor U \rfloor$ and return X_I.

  In other words, we select with replacement one of the data point.
- Better yet, use the function "sample" in R.

Nonparametric bootstrap

Example:

- Suppose that we have data $x_1, \ldots, x_n$, that we assume to be realizations of a random sample $X_1, \ldots, X_n$ iid f and we want to estimate $\theta = e^{-\mu}$, where μ is the population mean ($\mu = \int xf(x)dx$).
- A very natural estimate of $e^{-\mu}$ is $\hat{\theta} = e^{-\bar{X}}$, where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$.
- Clearly, this estimator is biased (why?). What is the bias of $\hat{\theta}$?

Nonparametric bootstrap

Example (continued):

- Since we do not assume a particular model $\{f_\theta\}$ here, the idea of bootstrap proceeds as follows.
- First we use the data $x_1, \ldots, x_n$ to compute the empirical cdf $\hat{F}$, which we view as an estimate of the true cdf F.
- We draw several random samples $(X_1^{(i)}, \ldots, X_n^{(i)})$ iid $\hat{F}$, for $i = 1, \ldots, B$.

Nonparametric bootstrap

Example (continued):

- For each one of them, we compute the estimator $\hat{\theta}^{(i)} = e^{-\bar{X}^{(i)}}$, $\bar{X}^{(i)} = n^{-1} \sum_{k=1}^{n} X_k^{(i)}$.
- We can now estimate the bias of the estimator by naturally comparing each $\hat{\theta}^{(i)}$ to $\hat{\theta}$.
- To carry the details, let's assume that f is the Gamma distrbution $Gamma(2, 5)$. In R, the nonparam. bootstrap algorithm becomes:

Nonparametric bootstrap

```
n=30 #sample size
X=rgamma(n,2,rate=5)# The data set
hatEst=exp(-mean(X)) #estimate
B=200 #Number of Bootstrap replication
Vec=numeric(B)
for (i in 1:B){
   U=runif(n,1,n+1); S=X[floor(U)]
   Vec[i]=exp(-mean(S))
}
hist(Vec-hatEst,30,main='hist.',xlab='',ylab='')
mean(Vec-hatEst) #bootstrap estimate of the bias
[1] 0.0006031303
```