The Bootstrap

Advanced Methods for Data Analysis (36-402/36-608)

Spring 2014

1 The bootstrap

1.1 Basic idea

- The *bootstrap* is one of the most general and the most widely used tools to estimate measures of uncertainty associated with a given statistical method. Some common bootstrap applications are: estimating the bias or variance of a particular statistical estimator, or constructing approximate confidence intervals for parameters of interest

- Basic setup: suppose that we have independent samples $z_1, \ldots z_n \sim P_\theta$. The subscript θ emphasizes the fact that θ is some parameter of interest, defined at the population level. E.g., this could be the mean of the distribution, the variance, or something more complicated. Let $\hat{\theta}$ be an estimate for θ that we compute from the samples $z_1, \ldots z_n$

- We may be interested knowing $\mathrm{Var}(\hat{\theta})$, the variance of our statistic $\hat{\theta}$. If we had access to P_θ, then we could just draw another fresh n samples, recompute the statistic, and repeat; after doing this, say, 1000 times, we would have computed 1000 statistics, and could just take the sample variance of these statistics

  However, of course, we generally don't have access to P_θ. The idea behind the bootstrap is to use the observed samples $z_1, \ldots z_n$ to generate n "new" samples, as if they came from P_θ. In particular, denoting the new samples by $\tilde{z}_1, \ldots \tilde{z}_n$, we draw these according to

  $$\tilde{z}_j \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\{z_1, \ldots z_n\}, \quad i = 1, \ldots n, \tag{1}$$

  in other words, each $\tilde{z}_j$ is independent and drawn uniformly among $z_1, \ldots z_n$. This is called *sampling with replacement*, because in our new sample $\tilde{z}_1, \ldots \tilde{z}_n$ we could very well have repeated observations

  In fact, we can think of $\tilde{z}_1, \ldots \tilde{z}_n$ as coming from a distribution—it is just an independent sample of size n from the *empirical distribution function* over the original sample $z_1, \ldots z_n$. This is a discrete distribution, with probability mass $1/n$ at each of $z_1, \ldots z_n$

  Now that we have this bootstrap scheme, to get an estimate for $\mathrm{Var}(\hat{\theta})$, we can just draw a bootstrap sample, recompute our statistic, repeat many times, and finally compute the sample variance over the statistics, as we would have done with samples from P_θ directly. More details on this next

1.2 A running example

- It helps to look at a specific example; here is a nice one from Chapter 5 of the ISL textbook. Suppose that we have two random variables $X, Y \in \mathbb{R}$ which represent the yields of two financial assets. We will invest a fraction of our money θ in X, and the remaining fraction $1-\theta$ in Y. Our yield will hence be

$$\theta X + (1-\theta)Y.$$

Because this is a random quantity, we may want to choose θ to minimize the variance of our investment. One can show that the value of θ minimizing

$$\mathrm{Var}\big(\theta X + (1-\theta)Y\big),$$

is indeed

$$\theta = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}},$$

where $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$, $\sigma_{XY} = \mathrm{Cov}(X, Y)$

- Given n samples of past measurements $(x_1, y_1), \dots (x_n, y_n)$ for the returns, we can compute estimates $\hat{\sigma}_X^2$, $\hat{\sigma}_Y^2$, and $\hat{\sigma}_{XY}$, which are the sample variances and sample covariance. Our plug-in estimate for θ is therefore

$$\hat{\theta} = \frac{\hat{\sigma}_Y^2 - \hat{\sigma}_{XY}}{\hat{\sigma}_X^2 + \hat{\sigma}_Y^2 - 2\hat{\sigma}_{XY}}$$

- Note: even if we new a parametric form (say bivariate normal) for the joint distribution of X and Y, performing formal calculations involving $\hat{\theta}$ would be difficult, because of the presence of sample estimates (sample variances and covariance) in its numerator and denominator

1.3 Estimating standard errors

- So how could we estimate the standard error of our estimator $\hat{\theta}$? (This is just its standard deviation; we often call the standard deviation of an estimator its standard error.) Use the bootstrap! Let $z_1 = (x_1, y_1), \dots z_n = (x_n, y_n)$ to make the notation consistent with that used in the last section. Then pick a large number B, say $B = 1000$, and repeat for $b = 1, \dots B$:
  - draw a bootstrap sample $\tilde{z}_1^{(b)}, \dots \tilde{z}_n^{(b)}$ as in (1);
  - recompute the statistic $\tilde{\theta}^{(b)}$ on $\tilde{z}_1^{(b)}, \dots \tilde{z}_n^{(b)}$.

  Then we to estimate the standard error $\mathrm{SE}(\hat{\theta})$, we use

$$\mathrm{SE}(\hat{\theta}) \approx \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\tilde{\theta}^{(b)} - \frac{1}{B}\sum_{r=1}^{B}\tilde{\theta}^{(r)}\right)^2}, \tag{2}$$

  which is just the sample standard deviation of the bootstrap statistics $\tilde{\theta}^{(1)}, \dots \tilde{\theta}^{(B)}$

1.4 Estimating bias

- We can also use the bootstrap to estimate the bias of our estimator. That (2) is a reasonable approximation is more or less very intuitive, but the bias argument is not as obvious. The idea is to make the approximation

$$\mathbb{E}(\hat{\theta}) - \theta \approx \mathbb{E}(\tilde{\theta}) - \hat{\theta} \tag{3}$$

$$\approx \frac{1}{B}\sum_{b=1}^{B}\tilde{\theta}^{(b)} - \hat{\theta} \tag{4}$$

- Once we believe (3), the second approximation (4) clearly follows. But why should (3) be reasonable? It will remain a valid approximation as long as the distributions of $\hat{\theta} - \theta$ and $\tilde{\theta} - \hat{\theta}$ are close. This is weaker than saying that the distributions of $\hat{\theta}$ and $\tilde{\theta}$ should be close, or that $\mathbb{E}(\hat{\theta})$ and θ should be close

  More generally, you may consider (3) to be a reasonable approximation as long as $\hat{\theta} - \theta$ is (roughly) *pivotal*, meaning that its distribution does not depend on the unknown parameter θ

2 Bootstrap confidence intervals

2.1 Basic bootstrap confidence intervals

- An extremely useful application of the bootstrap is the construction of *confidence intervals*. Recall that a $(1-\alpha)$ confidence interval for θ, computed over $z_1, \ldots z_n$, is a random interval $[L, U]$ satisfying

$$\mathbb{P}(L \leq \theta \leq U) = 1 - \alpha.$$

  We stress again the lower and upper limits L and U are random (i.e., L and U depend on $z_1, \ldots z_n$), and it is this randomness that is being considered in the probability statement above—the underlying parameter θ itself is fixed

- The *basic bootstrap confidence interval* for θ computes the bootstrap statistics $\tilde{\theta}^{(1)}, \ldots \tilde{\theta}^{(B)}$ as above, and then approximates the distribution of $\hat{\theta} - \theta$ by $\tilde{\theta} - \hat{\theta}$

  I.e., we compute the $\alpha/2$ and $1-\alpha/2$ quantiles of $\tilde{\theta}^{(1)}, \ldots \tilde{\theta}^{(B)}$, call them $q_{\alpha/2}$ and $q_{1-\alpha/2}$, and then argue

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\big(q_{\alpha/2} \leq \tilde{\theta} \leq q_{1-\alpha/2}\big) \\ &= \mathbb{P}\big(q_{\alpha/2} - \hat{\theta} \leq \tilde{\theta} - \hat{\theta} \leq q_{1-\alpha/2} - \hat{\theta}\big) \\ &\approx \mathbb{P}\big(q_{\alpha/2} - \hat{\theta} \leq \hat{\theta} - \theta \leq q_{1-\alpha/2} - \hat{\theta}\big) \\ &= \mathbb{P}\big(q_{\alpha/2} - 2\hat{\theta} \leq -\theta \leq q_{1-\alpha/2} - 2\hat{\theta}\big) \\ &= \mathbb{P}\big(2\hat{\theta} - q_{1-\alpha/2} \leq -\theta \leq 2\hat{\theta} - q_{\alpha/2}\big). \end{aligned}$$

  In other words, we use $[L, U] = [2\hat{\theta} - q_{1-\alpha/2}, 2\hat{\theta} - q_{\alpha/2}]$ as an approximate $(1-\alpha)$ confidence interval for θ

2.2 Studentized bootstrap confidence intervals

- If the distributions $\hat{\theta} - \theta$ and $\tilde{\theta} - \hat{\theta}$ are not close, then the basic bootstrap confidence interval can be inaccurate

- But even in this case, the distributions of $(\hat{\theta} - \theta)/\widehat{\mathrm{SE}}(\hat{\theta})$ and $(\tilde{\theta} - \hat{\theta})/\widehat{\mathrm{SE}}(\tilde{\theta})$ could be close, where $\widehat{\mathrm{SE}}(\cdot)$ denote estimated standard errors. Hence we could use what are called *studentized bootstrap confidence intervals*

- In this construction, we actually need two levels of boostrapping, an outer and inner bootstrap. We repeat, for $b = 1, \ldots B$:
  - draw a bootstrap sample $\tilde{z}_1^{(b)}, \ldots \tilde{z}_n^{(b)}$ from $z_1, \ldots z_n$;
  - recompute the statistic $\tilde{\theta}^{(b)}$ on $\tilde{z}_1^{(b)}, \ldots \tilde{z}_n^{(b)}$;
  - repeat, for $m = 1, \ldots M$:
    * draw a bootstrap sample $\tilde{z}_1^{(b,m)}, \ldots \tilde{z}_n^{(b,m)}$ from $\tilde{z}_1^{(b)}, \ldots \tilde{z}_n^{(b)}$;

    * recompute the statistic $\tilde{\theta}^{(b,m)}$ on $\tilde{z}_1^{(b,m)}, \ldots \tilde{z}_n^{(b,m)}$;
  - compute the sample standard deviation $\tilde{s}^{(b)}$ of $\tilde{\theta}^{(b,1)}, \ldots \tilde{\theta}^{(b,M)}$.

  At the end, we compute the sample standard deviation s of $\tilde{\theta}^{(1)}, \ldots \tilde{\theta}^{(B)}$. We also compute the $\alpha/2$ and $1-\alpha/2$ quantiles, call them $q_{\alpha/2}$ and $q_{1-\alpha/2}$, of

$$\frac{\tilde{\theta}^{(b)} - \hat{\theta}}{\tilde{s}^{(b)}}, \quad b = 1, \ldots B$$

- Now we make the argument

$$\begin{aligned} 1 - \alpha &= \mathbb{P}\left(q_{\alpha/2} \le \frac{\tilde{\theta} - \hat{\theta}}{\tilde{s}} \le q_{1-\alpha/2} \right) \\ &\approx \mathbb{P}\left(q_{\alpha/2} \le \frac{\hat{\theta} - \theta}{s} \le q_{1-\alpha/2} \right) \\ &= \mathbb{P}\big(s q_{\alpha/2} \le \hat{\theta} - \theta \le s q_{1-\alpha/2} \big) \\ &= \mathbb{P}\big(s q_{\alpha/2} - \hat{\theta} \le -\theta \le s q_{1-\alpha/2} - \hat{\theta} \big) \\ &= \mathbb{P}\big(\hat{\theta} - s q_{1-\alpha/2} \le \theta \le 2\hat{\theta} - q_{\alpha/2} \big). \end{aligned}$$

  Therefore we use

$$[L, U] = \big[\hat{\theta} - s q_{1-\alpha/2}, \hat{\theta} - s q_{\alpha/2} \big]$$

  as an approximate $(1-\alpha)$ confidence interval for θ. The advantage of this over the basic bootstrap confidence interval is that it can be more accurate; the disadvantage is that it is much more computationally demanding!