Consistency of Estimators

Guy Lebanon

May 1, 2006

It is satisfactory to know that an estimator $\hat{\theta}$ will perform better and better as we obtain more examples. If at the limit $n \to \infty$ the estimator tend to be always right (or at least arbitrarily close to the target), it is said to be consistent. This notion is equivalent to convergence in probability defined below.

Definition 1. *Let $Z_1, Z_2, \ldots$ be iid RVs and Z be a RVs. We say that Z_n converge in probability to Z, denoted $Z_n \to^p Z$ if*

$$\lim_{n\to\infty} P(|Z_n - Z| \leq \epsilon) = 1 \qquad \forall \epsilon > 0.$$

Consistency is defined as above, but with the target θ being a deterministic value, or a RV that equals θ with probability 1.

Definition 2. *Let $X_1, X_2, \ldots$ be a sequence of iid RVs drawn from a distribution with parameter θ and $\hat{\theta}$ an estimator for θ. We say that $\hat{\theta}$ is consistent as an estimator of θ if $\hat{\theta} \to^p \theta$ or*

$$\lim_n P(|\hat{\theta}(X_1, \ldots, X_n) - \theta| \leq \epsilon) = 1 \qquad \forall \epsilon > 0.$$

Consistency is a relatively weak property and is considered necessary of all reasonable estimators. This is in contrast to optimality properties such as efficiency which state that the estimator is "best".

Consistency of $\hat{\theta}$ can be shown in several ways which we describe below. The first way is using the law of large numbers (LLN) which states that an average $\frac{1}{n}\sum f(X_i)$ converges in probability to its expectation $\mathsf{E}(f(X_i))$. This can be used to show that $\bar{X}$ is consistent for $\mathsf{E}(X)$ and $\frac{1}{n}\sum X_i^k$ is consistent for $\mathsf{E}(X^k)$.

The second way is using the following theorem.

Theorem 1. *An unbiased estimator $\hat{\theta}$ is consistent if* $\lim_n \mathsf{Var}(\hat{\theta}(X_1, \ldots, X_n)) = 0$.

Proof. Since $\hat{\theta}$ is unbiased, we have using Chebyshev's inequality $P(|\hat{\theta} - \theta| > \epsilon) \leq \mathsf{Var}(\hat{\theta})/\epsilon^2$. For any $\epsilon > 0$, if $\mathsf{Var}(\hat{\theta}(X_1, \ldots, X_n)) = 0$, then $P(|\hat{\theta} - \theta| > \epsilon) \to 0$ as well. □

The third way of proving consistency is by breaking the estimator into smaller components, finding the limits of the components, and then piecing the limits together. This is described in the following theorem and example.

Theorem 2. *Let $\hat{\theta} \to_p \theta$ and $\hat{\eta} \to_p \eta$. Then*

1. $\hat{\theta} + \hat{\eta} \to_p \theta + \eta$.
2. $\hat{\theta}\hat{\eta} \to_p \theta\eta$.
3. $\hat{\theta}/\hat{\eta} \to_p \theta/\eta$ *if* $\eta \neq 0$.
4. $\hat{\theta} \to_p \theta \Rightarrow g(\hat{\theta}) \to_p g(\theta)$ *for any real valued function that is continuous at θ.*

The above theorem can be used to prove that S^2 is a consistent estimator of $\mathsf{Var}(X_i)$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) = \frac{n}{n-1}\left(\frac{1}{n}\sum X_i^2 - \bar{X}^2\right).$$

By the LLN $\bar{X} \to_p \mathsf{E}(X)$ and $\frac{1}{n}\sum X_i^2 \to_p \mathsf{E}(X^2)$ and by adding the above theorem we have $\bar{X}^2 \to_p \mathsf{E}(X)^2$. Using the theorem again we have $\left(\frac{1}{n}\sum X_i^2 - \bar{X}^2\right) \to_p \mathsf{E}(X^2) - \mathsf{E}(X)^2 = \mathsf{Var}(X)$ and again gives us the desired result $S^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum X_i^2 - \bar{X}^2\right) \to_p 1 \cdot \mathsf{Var}(X)$.

The following theorem connects between convergence in probability and convergence in distribution.

Definition 3. *A sequence of RVs X_n converges in distribution to X if $F_{X_n}(r) \to F_X(r)$ for all r at which F_X is continuous.*

Convergence in distribution is the convergence concept described in the central limit theorem. In general, it is a weaker form of convergence than convergence in probability. The following theorem connects the two concepts.

Theorem 3 (Slutsky's theorem)**.** *If X_n converges in distribution to X and Y_n converges in probability to c then X_nY_n converges in distribution to cX and $X_n + Y_n$ converges in distribution to $c + X$.*

One of the applications of the above theorem is that it show that the studentized RV $\sqrt{n}\frac{\bar{X}-\mu}{S} = \sqrt{n}\frac{\sigma}{S}\frac{\bar{X}-\mu}{\sigma}$ converges in distribution to $N(0,1)$ RV, where we used the fact that

$$S^2 \to_p \sigma^2 \Rightarrow S \to_p \sigma \Rightarrow \sigma/S \to_p 1/1 = 1$$

and the central limit theorem. This is a very potent result as it can be used instead of the CLT when σ^2 is unknown.