Monte Carlo Tests

Suppose we have a null hypothesis H_0, represented by a completely specified model, and a test statistic T, such as a goodness-of-fit statistic, for which small values indicate departure from H_0.
Denote the observed data by y and let the observed value of T be $t = T(y)$. small $T \implies$ rejection arbitrarily chosen.

Monte Carlo Tests

To perform a pure significance test we need to know the distribution of T under H_0. This may be difficult or impossible to obtain analytically, but it may be possible to simulate from the model to produce m samples $y_1^*, \ldots, y_m^*$ and the corresponding values $t_1^*, \ldots, t_m^*$ of T. We could then estimate the distribution of T under H_0 by the empirical distribution of $(t_1^*, \ldots, t_m^*)$, or a smoothed version thereof. Informally, we estimate the critical point of a level α test by the 100αth percentile of $(t_1^*, \ldots, t_m^*)$.

Monte Carlo Tests

Monte Carlo tests are a related idea. If H_0 is true, we have $m+1$ values from the distribution of T, m by simulation and one by observation. Thus the probability that t is the kth smallest or smaller of $\{t, t_1^*, \ldots, t_m^*\}$ is $k/(m+1)$, ignoring ties. [Assume a continuous distribution for T.] If we choose k and m so that $\alpha = k/(m+1)$ is a conventional significance level (e.g. 0.01 or 0.05), we have a 'Monte Carlo test', which rejects H_0 if t is the kth smallest or smaller of the $m+1$ values. [Two sided versions of the Monte Carlo test follow an obvious construction].

Example

Consider n random points in $[0,1] \times [0,1]$. The null hypothesis is one of uniformity, to be tested against one of inhibition between points. Let $T = \#$ 'close' pairs of points. Small values of T reject H_0. A Monte Carlo test of the significance of t is performed by generating m samples of n points uniformly on $[0,1] \times [0,1]$. Notice that in performing the test we can stop when k samples t_i^* are smaller than t, for then we know we will not reject. Also, if $m - k + 1$ values exceed t, we will certainly reject and can stop.

Monte Carlo Tests

Let F be the distribution function of T under H_0.

A 'conventional test of size α', possible if F were known, rejects H_0 if

$$u = F(t) = P(T \leq t \mid H_0) < \alpha.$$

The Monte Carlo test rejects H_0 if

$$t < t_{(k)},$$

where $t_{(k)}$ is the kth order statistic of $(t_1^*, \ldots, t_m^*)$.

Monte Carlo Tests

So, the Monte Carlo test rejects H_0 with probability p, where

$$\begin{aligned} p &= P(r \text{ simulated values of } T \text{ are } \leq t,\ 0 \leq r \leq k-1\ ) \\ &= \sum_{r=0}^{k-1} P(r \text{ simulated values of } T \text{ are } \leq t) \\ &= \sum_{r=0}^{k-1} \binom{m}{r} u^r (1-u)^{m-r}. \qquad (u = F(t)) \end{aligned}$$

This should be interpreted as the proportion of times the Monte Carlo test will reject H_0 with $t = F^{-1}(u)$ as observation. Note: we are not assuming $t \sim F$ here, just using F^{-1} as a transformation, as we have $t_i^* \sim F_T$

Monte Carlo Tests

Compare the power function $\beta^{(m)}(\alpha)$ of the Monte Carlo test with the power $\beta(\alpha)$ of the conventional test. We expect, and it is true, that $\beta^{(m)} \leq \beta$, but how large can the power loss be? Power = 1-P(type II error) = P(reject | H_0 false)

The conventional test rejects H_0 if $T < F^{-1}(\alpha)$ and has power against an alternative denoted by F_θ of

$$\begin{aligned} \beta(\alpha) &= P\big(T < F^{-1}(\alpha) \mid T \sim F_\theta\big) \\ &= F_\theta\big(F^{-1}(\alpha)\big). \end{aligned}$$

Monte Carlo Tests

Similarly, the power of the Monte Carlo test against this alternative is

$$\begin{aligned}
\beta^{(m)}(\alpha) &= P(T < t_{(k)} \mid T \sim F_\theta, t_1^*, \ldots, t_m^* \sim F) \\
&= \int_{-\infty}^{\infty} \sum_{r=0}^{k-1} \binom{m}{r} F(t)^r [1 - F(t)]^{m-r} dF_\theta(t) \\
&= \int_{-\infty}^{\infty} \int_{F(t)}^{1} b(\alpha, m, \xi) d\xi dF_\theta(t) \\
&= \int_0^1 F_\theta\big(F^{-1}(\xi)\big) b(\alpha, m, \xi) d\xi \\
&= \int_0^1 \beta(\xi) b(\alpha, m, \xi) d\xi,
\end{aligned}$$

where $b(\alpha, m, \cdot)$ is the pdf of a beta distribution with parameters $\alpha(m+1)$ and $(1-\alpha)(m+1)$.

Monte Carlo Tests

If m is big, the Monte Carlo test is OK.

Avoid small values of k, which blur the critical region a lot and lose power. Since $\alpha = k/(m+1)$, this implies m is large. Usually $m = 99$ will be sufficient.

Resampling Techniques: The Jackknife

Suppose we have a statistic θ, with an estimator

$$\hat{\theta} = \hat{\theta}\left(X_1, \ldots, X_n\right).$$

Construct n distinct $(n-1)$-length samples from our original sample $X_1, \ldots, X_n$; do this by leaving out each datapoint in turn.

$$\underline{X}_{-i} = \underline{X}_{(i)} = \left(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\right), i = 1, \ldots, n$$

Use each sample $\underline{X}_{(i)}$ to obtain a different estimate $\hat{\theta}_{(i)} = \hat{\theta}\left(\underline{X}_{(i)}\right)$

Resampling Techniques: The Jackknife

Consider

$$\hat{\theta}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}.$$

with variance estimator

$$\widehat{\mathrm{V}}_{jack}\left[\hat{\theta}\right] = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2.$$

Resampling Techniques: Bootstrap Resampling

Suppose we have an i.i.d. sample from an unknown distribution, with cdf $F(\cdot)$. Given the data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F(x)$, construct an empirical estimate $\widehat{F}$ of the distribution function

$$\widehat{F}(x) = \frac{i}{n} \qquad X_{(i)} \leq x < X_{(i+1)}$$

where $X_{(i)}$ is the i^{th} order statistic. Sample B i.i.d. samples from this empirical estimate,

$$X_1^{*(b)}, \ldots, X_n^{*(b)} \overset{i.i.d.}{\sim} \widehat{F}, \qquad b = 1, \ldots, B.$$

Resampling Techniques: Bootstrap Resampling

We define the **Monte Carlo approximation of the bootstrap estimator** for the variance of

$$\widehat{\theta}^{*(b)} = \widehat{\theta}\left(X_1^{*(b)}, \ldots, X_n^{*(b)}\right), \qquad b = 1, \ldots, B.$$

to be

$$\widehat{\mathrm{Var}}^{*}_{boot}\left[\widehat{\theta}\right] = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\theta}^{*(b)} - \frac{1}{B}\sum_{k=1}^{B}\widehat{\theta}^{*(k)}\right)^2.$$

Resampling Techniques: Bootstrap Resampling

- This methodology extends flexibly to estimating any distributional characteristics of $\widehat{\theta}$, not just the variance!
- number of bootstrap samples is not restricted by the size of the original data sample, unlike with the Jackknife
- Monte Carlo sampling from $\widehat{F} \equiv$ resampling with replacement from the original data!
- Increasing B will reduce the Monte Carlo error, but not the error due to using $\hat{F}$ instead of F.