Testing Random-Number Generators

Goal: To ensure that the random number generator produces a random stream.

- Plot histograms
- Plot quantile-quantile plot
- Use other tests
- Passing a test is necessary but not sufficient
- Pass ≠ Good
- Fail ⇒ Bad
- New tests ⇒ Old generators fail the test
- Tests can be adapted for other distributions

Chi-Square Test

- Most commonly used test
- Can be used for any distribution
- Prepare a histogram of the observed data
- Compare observed frequencies with theoretical
  \( k \) = Number of cells
  \( o_i \) = Observed frequency for \( i \)th cell
  \( e_i \) = Expected frequency

\[
D = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}
\]

- \( D \approx \chi^2 \) distribution with \( k - 1 \) degrees of freedom.
- Compare \( D \) with \( \chi^2_{1-a, k-1} \). Pass with confidence \( a \) if \( D \) is less

Example 27.1

- 1000 random numbers with \( x_0 = 1 \)
- \( x^2_{30,0.05} = 14.68 \)
- Observed difference = 10.380
- Observed is Less ⇒ Accept IID U(0, 1)

Chi-Square for Other Distributions

- Errors in cells with a small \( e_i \) affect the chi-square statistic more
- Best when \( e_i \)'s are equal.

\[
\text{Use an equi-probable histogram with variable cell sizes}
\]

- Combine adjoining cells so that the new cell probabilities are approximately equal.
- The number of degrees of freedom should be reduced to \( k-r-1 \) (in place of \( k-1 \)), where \( r \) is the number of parameters estimated from the sample.
- Designed for discrete distributions and for large sample sizes only ⇒ Lower significance for finite sample sizes and continuous distributions
- If less than 5 observations, combine neighboring cells
Serial-Correlation Test

- Nonzero covariance ⇒ Dependence. The inverse is not true.
- \( R_k = \text{Autocovariance at lag } k = \text{Cov}(x_n, x_{n+k}) \)
  \[
  R_k = \frac{1}{n-k} \sum_{i=1}^{n-k} (U_i - \frac{1}{2})(U_{i+k} - \frac{1}{2})
  \]
- For large \( n \), \( R_k \) is normally distributed with a mean of zero and a variance of \( 1/(144(n-k)) \).
- 100(1-\( \alpha \))% confidence interval for the autocovariance is:
  \[
  R_k \pm z_{1-\alpha/2}/(12\sqrt{n-k})
  \]

For \( k = 1 \) Check if CI includes zero.
For \( k = 0 \), \( R_0 \) must be variance of the sequence. Expected to be \( 1/12 \) for IID \( U(0,1) \).

Example 27.3: Serial Correlation Test

- \( x_n = 7^{n} x_{n-1} \mod (2^{31} - 1) \)
- 10,000 random numbers with \( x_0 = 1 \):

<table>
<thead>
<tr>
<th>Lag</th>
<th>Autocovariance</th>
<th>St. Dev.</th>
<th>90% Confidence Interval</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.000038</td>
<td>0.000833</td>
<td>-0.001409</td>
<td>0.001333</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.001017</td>
<td>0.000833</td>
<td>-0.002388</td>
<td>0.000354</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.000489</td>
<td>0.000833</td>
<td>-0.001860</td>
<td>0.000882</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-0.000333</td>
<td>0.000834</td>
<td>-0.001404</td>
<td>0.001339</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-0.000531</td>
<td>0.000834</td>
<td>-0.001902</td>
<td>0.000840</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.001277</td>
<td>0.000834</td>
<td>-0.002648</td>
<td>0.000995</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-0.000385</td>
<td>0.000834</td>
<td>-0.001757</td>
<td>0.000986</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-0.000207</td>
<td>0.000834</td>
<td>-0.001579</td>
<td>0.001164</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.001001</td>
<td>0.000834</td>
<td>-0.000034</td>
<td>0.002403</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-0.000224</td>
<td>0.000834</td>
<td>-0.001595</td>
<td>0.001148</td>
<td></td>
</tr>
</tbody>
</table>

Example 27.3 (Cont)

- All confidence intervals include zero ⇒ All covariances are statistically insignificant at 90% confidence.

Serial Test

- Goal: To test for uniformity in two dimensions or higher.
- In two dimensions, divide the space between 0 and 1 into \( K^2 \) cells of equal area.
- Given \( \{x_1, x_2, ..., x_n\} \), use \( n/2 \) non-overlapping pairs \( (x_p, x_q) \), \( (x_r, x_s) \), ... and count the points in each of the \( K^2 \) cells.
- Expected= \( n/(2K^2) \) points in each cell.
- Use chi-square test to find deviation of the actual counts from the expected counts.
- The degrees of freedom in this case are \( K^2-1 \).
- For \( k \)-dimensions, use \( k \)-tuples of non-overlapping values.
- \( k \)-tuples must be non-overlapping.
- Overlapping ⇒ number of points in the cells are not independent chi-square test cannot be used.
- In visual check one can use overlapping or non-overlapping.
- In the spectral test overlapping tuples are used.
- Given \( n \) numbers, there are \( n-1 \) overlapping pairs, \( n/2 \) non-overlapping pairs.

Summary

1. Chi-square test is a one-dimensional test
   Designed for discrete distributions and large sample sizes
2. Serial correlation test for independence
3. \( k \)-dimensional uniformity = \( k \)-distributivity tested by spectral test or serial test