Eugenia Stoimenova
Two-sample rank tests based on exceeding observations

Applications of Mathematics, Vol. 52 (2007), No. 4, 345--352

Persistent URL: http://dml.cz/dmlcz/134680

Terms of use:
© Institute of Mathematics AS CR, 2007

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project DML-CZ: The Czech Digital Mathematics Library http://project.dml.cz
TWO-SAMPLE RANK TESTS BASED ON 
EXCEEDING OBSERVATIONS 

EUGENIA STOIMENOVA, Sofia 

(Received January 1, 2006, in revised version April 24, 2006) 

Abstract. Simple rank statistics are used to test that two samples come from the same 
distribution. Šidák’s $E$-test (Apl. Mat. 22 (1977), 166–175) is based on the number of 
observations from one sample that exceed all observations from the other sample. A similar 
test statistic is defined in Ann. Inst. Stat. Math. 52 (1970), 255–266. We study asymptotic 
behavior of the moments of both statistics.

Keywords: location problem, $E$-test statistic, $M$-test statistic

MSC 2000: 62G10, 62G20

1. Introduction

Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be random samples from distributions $F$ and $G$, 
respectively. We wish to test the hypothesis $H_0$ that $F$ and $G$ are identical against 
the alternative that $F(x) \geq G(x)$, with strict inequality for some $x$.

Rank statistics based on the number of observations from one sample which exceed 
all observations from the other sample give rise to quick and easy tests which are 
suitable for testing that two samples come from the same distribution. The most 
popular of tests of this type are Haga’s test [1] and Šidák’s $E$-test [4]. Some other 
tests based on exceeding observations are discussed in [2], [3], [5]. Recently a new 
test statistic, $M$-statistic, is proposed in [6] for the two sample problem. In this 
paper we study the asymptotic behavior of the moments of $M$-statistic and use the 
results for deriving analogous properties of $E$-statistic.

The tests based on exceeding observations can be described with reference to the 
same basic situation. The notation is adapted from Hájek and Šidák (1967). We 
define $A$ and $B'$ to be the number of observations among $X_1, \ldots, X_m$ larger than 
$max Y_j$, or smaller than $min Y_j$, respectively, and $A'$ and $B$ to be the number
of observations among $Y_1, \ldots, Y_n$ larger than $\max_{1 \leq i \leq m} X_i$, or smaller than $\min_{1 \leq i \leq m} X_i$, respectively. Clearly, only one of the numbers $A$ and $A'$ (or $B$ and $B'$) is positive, while the other must be zero.

With this notation, the $E$-statistic for testing $H_0$ is defined by

$$E = \min(A, B) - \min(A', B'),$$

and the $M$-statistic is defined by

$$M = \max\{m - A, n - B\}.$$  

These statistics are non-linear rank statistics and are not asymptotically normally distributed. In Section 2 we prove the asymptotic distribution of $M$ statistic for largest values. Some intermediate results are given there. In Section 3 we derive the asymptotic distribution of the mean and the variance of $M$ under the null hypothesis.

### 2. Definitions and presentations

Suppose that the notation is chosen so that $m < n$. The exact distributions of $E$ and $M$ statistics under $H_0$ are presented in the corresponding papers as follows:

$$P(E \geq k) = P(E \leq -k) = \binom{m + n}{m}^{-1} \binom{m + n - 2k}{m - k}, \quad k = 1, \ldots, m,$$

$$P(M = k) = \begin{cases} \binom{m + n}{m}^{-1} \binom{2k - 2}{k - 1} \frac{3k - 2}{k}, & \text{for } 1 \leq k \leq m; \\ \binom{m + n}{m}^{-1} \binom{m + k - 1}{m - 1}, & \text{for } m < k \leq n. \end{cases} \quad (1)$$

For small sample sizes the above distributions are easily enumerated. For large sample sizes the vast majority of the mass is above $\min(m, n)$. The limit distribution of the upper tail of $E$-statistic is obtained in [4]. The next theorem gives the limit distribution of the largest values of $M$-statistic.

#### 2.1. Limit distribution of the $M$-statistic

**Theorem 1.** Let $m, n \to \infty$ and $m/n \to \lambda$ $(0 < \lambda < 1)$. Then for $0 \leq k \leq n - m - 1$

$$P(M = n - k) \longrightarrow \frac{\lambda}{1 + \lambda} \left(\frac{1}{1 + \lambda}\right)^k.$$
Proof. The probability $P(M = n - k)$ in (1) can be expressed

$$P(M = n - k) = \frac{m!n!(m - k - 1)!}{(m + n)!(m - 1)!} = \frac{mn(n - 1)\ldots(n - k + 1)}{(m + n)(m + n - 1)\ldots(m + n - k)}$$

which obviously tends to $\lambda(1 + \lambda)^{-1}(1 + \lambda)^{-k}$ as $m, n \to \infty$ and $m/n \to \lambda$. □

Corollary 1. The probability mass concentrated in the maximum value of $M$ is asymptotically equivalent to $\lambda(1 + \lambda)^{-1}$.

The approximate number of points in the lower tail $P\{M \leq k\} \leq \alpha$ can be calculated using Theorem 1. The approximation depends on the ratios of $m$ and $n$ and on the size of the second sample $n$ as well.

The first columns in Tabs. 1 and 2 contain different ratios of $m$ and $n$ ($0 < \lambda < 1$), and the first rows contain some large values for $n$. Similar tables are calculated in [6] using the exact distribution for $n = 1, \ldots, 25$; $m = 1, \ldots, n$; $\alpha = 0.01, 0.05$. A comparison between the exact distribution and the approximate values can be seen for $n = 25$.

<table>
<thead>
<tr>
<th>$\lambda$ \ $n$</th>
<th>25</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>150</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>*</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>120</td>
<td>170</td>
</tr>
<tr>
<td>0.14</td>
<td>3</td>
<td>18</td>
<td>23</td>
<td>28</td>
<td>33</td>
<td>38</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>73</td>
<td>83</td>
<td>103</td>
<td>183</td>
</tr>
<tr>
<td>0.18</td>
<td>8</td>
<td>23</td>
<td>28</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>68</td>
<td>78</td>
<td>88</td>
<td>98</td>
<td>128</td>
<td>178</td>
</tr>
<tr>
<td>0.22</td>
<td>11</td>
<td>26</td>
<td>31</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>76</td>
<td>86</td>
<td>96</td>
<td>106</td>
<td>136</td>
<td>186</td>
</tr>
<tr>
<td>0.26</td>
<td>13</td>
<td>28</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>68</td>
<td>78</td>
<td>88</td>
<td>98</td>
<td>108</td>
<td>138</td>
<td>188</td>
</tr>
<tr>
<td>0.30</td>
<td>15</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
<td>140</td>
<td>190</td>
</tr>
<tr>
<td>0.34</td>
<td>16</td>
<td>31</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>81</td>
<td>91</td>
<td>101</td>
<td>111</td>
<td>141</td>
<td>191</td>
</tr>
<tr>
<td>0.38</td>
<td>17</td>
<td>32</td>
<td>37</td>
<td>42</td>
<td>47</td>
<td>52</td>
<td>57</td>
<td>62</td>
<td>67</td>
<td>72</td>
<td>82</td>
<td>92</td>
<td>102</td>
<td>112</td>
<td>142</td>
<td>192</td>
</tr>
<tr>
<td>0.42</td>
<td>17</td>
<td>32</td>
<td>37</td>
<td>42</td>
<td>47</td>
<td>52</td>
<td>57</td>
<td>62</td>
<td>67</td>
<td>72</td>
<td>82</td>
<td>92</td>
<td>102</td>
<td>112</td>
<td>142</td>
<td>192</td>
</tr>
<tr>
<td>0.46</td>
<td>18</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>68</td>
<td>73</td>
<td>83</td>
<td>93</td>
<td>103</td>
<td>113</td>
<td>143</td>
<td>193</td>
</tr>
<tr>
<td>0.50</td>
<td>19</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>74</td>
<td>84</td>
<td>94</td>
<td>104</td>
<td>114</td>
<td>144</td>
<td>194</td>
</tr>
<tr>
<td>0.54</td>
<td>19</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>74</td>
<td>84</td>
<td>94</td>
<td>104</td>
<td>114</td>
<td>144</td>
<td>194</td>
</tr>
<tr>
<td>0.58</td>
<td>19</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>74</td>
<td>84</td>
<td>94</td>
<td>104</td>
<td>114</td>
<td>144</td>
<td>194</td>
</tr>
<tr>
<td>0.62</td>
<td>20</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>85</td>
<td>95</td>
<td>105</td>
<td>115</td>
<td>145</td>
<td>195</td>
</tr>
<tr>
<td>0.66</td>
<td>20</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>85</td>
<td>95</td>
<td>105</td>
<td>115</td>
<td>145</td>
<td>195</td>
</tr>
<tr>
<td>0.70</td>
<td>20</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>85</td>
<td>95</td>
<td>105</td>
<td>115</td>
<td>145</td>
<td>195</td>
</tr>
<tr>
<td>0.74</td>
<td>21</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>76</td>
<td>86</td>
<td>96</td>
<td>106</td>
<td>116</td>
<td>146</td>
<td>196</td>
</tr>
<tr>
<td>0.78</td>
<td>21</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>76</td>
<td>86</td>
<td>96</td>
<td>106</td>
<td>116</td>
<td>146</td>
<td>196</td>
</tr>
<tr>
<td>0.82</td>
<td>21</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>76</td>
<td>86</td>
<td>96</td>
<td>106</td>
<td>116</td>
<td>146</td>
<td>196</td>
</tr>
<tr>
<td>0.86</td>
<td>21</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>76</td>
<td>86</td>
<td>96</td>
<td>106</td>
<td>116</td>
<td>146</td>
<td>196</td>
</tr>
<tr>
<td>0.90</td>
<td>21</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>76</td>
<td>86</td>
<td>96</td>
<td>106</td>
<td>116</td>
<td>146</td>
<td>196</td>
</tr>
<tr>
<td>0.94</td>
<td>21</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>76</td>
<td>86</td>
<td>96</td>
<td>106</td>
<td>116</td>
<td>146</td>
<td>196</td>
</tr>
<tr>
<td>0.98</td>
<td>22</td>
<td>37</td>
<td>42</td>
<td>47</td>
<td>52</td>
<td>57</td>
<td>62</td>
<td>67</td>
<td>72</td>
<td>77</td>
<td>87</td>
<td>97</td>
<td>107</td>
<td>117</td>
<td>147</td>
<td>197</td>
</tr>
</tbody>
</table>

Table 1. Approximate number of points in the lower tail $P\{M \leq k\} \leq 0.05$ using Theorem 1.
<table>
<thead>
<tr>
<th>λ \ n</th>
<th>25</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
<th>150</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>*</td>
<td>*</td>
<td>3</td>
<td>8</td>
<td>13</td>
<td>18</td>
<td>23</td>
<td>28</td>
<td>33</td>
<td>43</td>
<td>53</td>
<td>63</td>
<td>73</td>
<td>103</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>0.14</td>
<td>*</td>
<td>6</td>
<td>11</td>
<td>16</td>
<td>21</td>
<td>26</td>
<td>31</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>56</td>
<td>66</td>
<td>76</td>
<td>116</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>0.18</td>
<td>*</td>
<td>13</td>
<td>18</td>
<td>23</td>
<td>28</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>63</td>
<td>73</td>
<td>83</td>
<td>123</td>
<td>173</td>
<td></td>
</tr>
<tr>
<td>0.22</td>
<td>3</td>
<td>18</td>
<td>23</td>
<td>28</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>68</td>
<td>78</td>
<td>88</td>
<td>128</td>
<td>178</td>
<td></td>
</tr>
<tr>
<td>0.26</td>
<td>6</td>
<td>21</td>
<td>26</td>
<td>31</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>71</td>
<td>81</td>
<td>91</td>
<td>131</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>8</td>
<td>23</td>
<td>28</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>73</td>
<td>83</td>
<td>93</td>
<td>133</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>0.34</td>
<td>10</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>75</td>
<td>85</td>
<td>95</td>
<td>135</td>
<td>185</td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>12</td>
<td>27</td>
<td>32</td>
<td>37</td>
<td>42</td>
<td>47</td>
<td>52</td>
<td>57</td>
<td>62</td>
<td>67</td>
<td>77</td>
<td>87</td>
<td>97</td>
<td>137</td>
<td>187</td>
<td></td>
</tr>
<tr>
<td>0.42</td>
<td>13</td>
<td>28</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>68</td>
<td>78</td>
<td>88</td>
<td>98</td>
<td>138</td>
<td>188</td>
<td></td>
</tr>
<tr>
<td>0.46</td>
<td>14</td>
<td>29</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>79</td>
<td>89</td>
<td>99</td>
<td>139</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>15</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>140</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>0.54</td>
<td>15</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>140</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>0.58</td>
<td>16</td>
<td>31</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>81</td>
<td>91</td>
<td>101</td>
<td>141</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>0.62</td>
<td>16</td>
<td>31</td>
<td>36</td>
<td>41</td>
<td>46</td>
<td>51</td>
<td>56</td>
<td>61</td>
<td>66</td>
<td>71</td>
<td>81</td>
<td>91</td>
<td>101</td>
<td>141</td>
<td>191</td>
<td></td>
</tr>
<tr>
<td>0.66</td>
<td>17</td>
<td>32</td>
<td>37</td>
<td>42</td>
<td>47</td>
<td>52</td>
<td>57</td>
<td>62</td>
<td>67</td>
<td>72</td>
<td>82</td>
<td>92</td>
<td>102</td>
<td>142</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>17</td>
<td>32</td>
<td>37</td>
<td>42</td>
<td>47</td>
<td>52</td>
<td>57</td>
<td>62</td>
<td>67</td>
<td>72</td>
<td>82</td>
<td>92</td>
<td>102</td>
<td>142</td>
<td>192</td>
<td></td>
</tr>
<tr>
<td>0.74</td>
<td>18</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>68</td>
<td>73</td>
<td>83</td>
<td>93</td>
<td>103</td>
<td>143</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>0.78</td>
<td>18</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>68</td>
<td>73</td>
<td>83</td>
<td>93</td>
<td>103</td>
<td>143</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>0.82</td>
<td>18</td>
<td>33</td>
<td>38</td>
<td>43</td>
<td>48</td>
<td>53</td>
<td>58</td>
<td>63</td>
<td>68</td>
<td>73</td>
<td>83</td>
<td>93</td>
<td>103</td>
<td>143</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>0.86</td>
<td>19</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>74</td>
<td>84</td>
<td>94</td>
<td>104</td>
<td>144</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>19</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>74</td>
<td>84</td>
<td>94</td>
<td>104</td>
<td>144</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>0.94</td>
<td>19</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>74</td>
<td>84</td>
<td>94</td>
<td>104</td>
<td>144</td>
<td>194</td>
<td></td>
</tr>
<tr>
<td>0.98</td>
<td>19</td>
<td>34</td>
<td>39</td>
<td>44</td>
<td>49</td>
<td>54</td>
<td>59</td>
<td>64</td>
<td>69</td>
<td>74</td>
<td>84</td>
<td>94</td>
<td>104</td>
<td>144</td>
<td>194</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Approximate number of points in the lower tail $P\{M \leq k\} \leq 0.01$ using Theorem 1.

Since for large sample sizes the vast majority of the mass is above $\min(m, n)$, Theorem 1 cannot be used for equal sample sizes. Analogous approximation for $m = n$ gives

$$P(M = m - k) \longrightarrow \frac{3}{4} \left(\frac{1}{4}\right)^k$$

as $m \to \infty$ and $k = 0, \ldots, m - 1$.

Here and further we suppose that $m = n$. Then the expectations of $M_n$ and $M_n^2$, respectively, under $H_0$ are represented as follows:

(2) \hspace{1cm} \mathbb{E}(M_n) = n - \left(\frac{2n}{n}\right)^{-1} \sum_{k=0}^{n-1} \left(\frac{2k}{k}\right),

(3) \hspace{1cm} \mathbb{E}(M_n^2) = n(n - 1) - \left(\frac{2n}{n}\right)^{-1} \sum_{k=0}^{n-1} \left(\frac{2k}{k}\right)(k + 1).$

The proof of (2) is via direct comparison with $\mathbb{E}M_n$ computed from the distribution (1) and defined by

$$\mathbb{E}(M_n) = \sum_{k=1}^{n} \left(\frac{2n}{n}\right)^{-1} \left(\frac{2k-2}{k-1}\right) \frac{3k-2}{k}.$$
Straightforward induction on $n$ gives the result. The proof of (3) is via direct comparison with $\mathbb{E}(M_n^2)$ and induction on $n$.

### 2.2. Some intermediate results

**Proposition 1.** Define the sequences

\[
q_n = \frac{\mathbb{E}(M_n)}{n},
\]

\[
q_n' = \mathbb{E}(M_n) - n,
\]

and

\[
q_n'' = n\left[\mathbb{E}(M_n) - n + \frac{1}{3}\right], \quad n \geq 1.
\]

1) Then $q_n$, $q_n'$ and $q_n''$ satisfy the recurrence relations

\[
q_{n+1} = \frac{n}{4n+2}q_n + \frac{3n+1}{4n+2},
\]

\[
q_{n+1}' = \frac{n+1}{4n+2}q_n' - \frac{n+1}{4n+2},
\]

and

\[
q_{n+1}'' = \frac{(n+1)^2}{n(4n+2)}q_n'' - \frac{n+1}{3(2n+1)}.
\]

2) We claim that $q_n$, $q_n'$ and $q_n''$ converge to finite limits as $n \to \infty$, and

\[
\lim_{n \to \infty} q_n = 1, \quad \lim_{n \to \infty} q_n' = -\frac{1}{3}, \quad \text{and} \quad \lim_{n \to \infty} q_n'' = -\frac{2}{9}.
\]

**Proof.** 1) The recurrence relations (4)–(6) follow immediately from (2) and simple algebra.

2) The proof of convergence of $q_n$, $q_n'$ and $q_n''$, which we omit, is analogous to the proof for the sequence $t_n$ in Theorem 2. The limits then follow from the recurrence relations. $\square$

**Proposition 2.** Define the sequences

\[
y_n = \frac{\mathbb{E}(M_n^2)}{n},
\]

\[
y_n' = \frac{1}{n}\left[\mathbb{E}(M_n^2) - n^2\right],
\]

and

\[
y_n'' = \left[\mathbb{E}(M_n^2) - n^2 + \frac{2}{3}n\right], \quad n \geq 1.
\]

349
1) Then $y_n$, $y'_n$ and $y''_n$ satisfy the recurrence relations

$$y_{n+1} = \frac{n}{4n+2}y_n + \frac{(3n+1)(n+1)}{4n+2},$$

(8)

$$y'_{n+1} = \frac{n}{4n+2}y'_n - \frac{n+1}{4n+2},$$

(9)

and

$$y''_{n+1} = \frac{n+1}{2n+1}y''_n - \frac{n+1}{3(2n+1)},$$

(10)

2) We claim that $y_n$, $y'_n$ and $y''_n$ converge to finite limits as $n \to \infty$, and

$$\lim_{n \to \infty} y_n = 1, \quad \lim_{n \to \infty} y'_n = -\frac{2}{3}, \quad \text{and} \quad \lim_{n \to \infty} y''_n = -\frac{2}{9}.$$  

Proof. 1) The recurrence relations (8)–(10) follow from (3) and simple algebra.  
2) The proof of convergence of $y_n$, $y'_n$ and $y''_n$ is analogous to the proof for $t_n$ in Theorem 2. The limits (11) follow from the recurrence relations. \[ \square \]

3. Asymptotic behavior of the moments

3.1. Expectation of the $M$-statistic

As a consequence of (7) we obtain the asymptotic behavior of the expectation of the $M$-statistic.

Theorem 2. As $n \to \infty$

$$\mathbb{E}(M_n) = n - \frac{1}{3} - \frac{2}{9n} - \frac{2}{9n^2}(1 + o(1)).$$  

Proof. Define $t_n = \mathbb{E}(M_n) - n + \frac{1}{3} + \frac{2}{9}n^{-1}$. Utilizing (2) it is easy to check that $t_n$ satisfies the recurrence relation

$$t_n = \frac{n}{4n-2}t_{n-1} - \frac{3n-2}{9(n-1)n(2n-1)}.$$  

(13)

Now, substitute $t_{n-1}$ by its recurrence relation (13), and repeat substitution until $t_1$ is reached. Straightforward manipulation of the sums gives

$$t_n = \left(\frac{2n}{n}\right)^{-1}2t_1 - \left(\frac{2n}{n}\right)^{-1}\sum_{k=1}^{n-1} \left(\frac{2k+2}{k+1}\right)\frac{3k+1}{9k(k+1)(2k+1)}.$$  

(14)

It is easily seen from (2) that $t_1 = \frac{1}{18}$.  

350
The presentation (14) of $t_n$ implies that $\frac{1}{9}$ is an upper bound for the sequence $t_n$. To prove the convergence of $t_n$ it is sufficient to show that it is increasing with $n$. That is,

$$
\begin{align*}
t_{n+1} \geq t_n & \iff \left(\frac{2n + 2}{n + 1}\right)^{-1} \frac{1}{9} - \left(\frac{2n + 2}{n + 1}\right)^{-1} \sum_{k=1}^{n} \left(\frac{2k + 2}{k + 1}\right) \frac{3k + 1}{9k(k + 1)(2k + 1)} \\
& \geq \left(\frac{2n}{n}\right)^{-1} \frac{1}{9} - \left(\frac{2n}{n}\right)^{-1} \sum_{k=1}^{n-1} \left(\frac{2k + 2}{k + 1}\right) \frac{3k + 1}{9k(k + 1)(2k + 1)} \\
& \iff \sum_{k=1}^{n-1} \left(\frac{2k + 2}{k + 1}\right) \frac{3k + 1}{9k(k + 1)(2k + 1)} \geq \left(\frac{2n + 2}{n + 1}\right) \frac{1}{9n(2n + 1)}.
\end{align*}
$$

The last inequality follows by a simple induction on $n$. Thus $t_n$ is increasing and hence convergent sequence.

Now, from the recurrence relation (13) we have

$$
\lim_{n \to \infty} t_n = \lim_{n \to \infty} \frac{n}{4n - 2} \lim_{n \to \infty} t_{n-1} - \lim_{n \to \infty} \frac{3n - 2}{n(2n - 1)}.
$$

Therefore, $\lim_{n \to \infty} t_n = -\frac{2}{9} \lim_{n \to \infty} 1/n^2 = 0$. □

### 3.2. Variance of the $M$-statistic

As a consequence of the results (11) we obtain the asymptotic behavior of the variance of the $M$-statistic.

**Theorem 3.** As $n \to \infty$

$$
\text{var}(M_n) = \frac{1}{9} + \frac{1}{9n}(1 + o(1)).
$$

**Proof.** Using the definition of the sequence $y''_n$ and the limits (11) we have

$$
\mathbb{E}(M_n^2) = n^2 - \frac{2}{3} n - \frac{2}{9}(1 + o(1)).
$$

Then from Proposition 1 and Proposition 2 we have

$$
\text{var}(M_n) = \mathbb{E}(M_n^2) - (\mathbb{E}(M_n))^2
\begin{align*}
&= n^2 - \frac{2}{3} n - \frac{2}{9} - \left(n - \frac{1}{3} - \frac{2}{9n}\right)^2 + O\left(\frac{1}{n^2}\right) \\
&= \frac{1}{9} + O\left(\frac{1}{n^2}\right).
\end{align*}
$$

□
3.3. Moments of the $E$-statistic

It is not hard to relate the moments of the distribution of the $E$-statistic to those of $M$-statistic. Note that the distribution of the $E$-statistic is symmetric about 0 so its expectation under $H_0$ is 0.

For the second moment we have

$$
\mathbb{E}(E_n^2) = 2 \sum_{k=1}^{n-1} \binom{2n}{n}^{-1} \left[ \binom{2n}{n-k} - \binom{2n-2k-2}{n-k-1} \right] k^2 + 2 \binom{2n}{n}^{-1} n^2
$$

Further,

$$
\mathbb{E}(E_{n+1}^2) = \frac{n+1}{4n+2} \left[ 2 \binom{2n}{n}^{-1} \sum_{k=0}^{n-1} \binom{2k}{k} (2n-2k-1) + 4 \binom{2n}{n}^{-1} \sum_{k=0}^{n-1} \binom{2k}{k} + 2 \right].
$$

Using (2) it follows that the presentation (16) of $\mathbb{E}(E_n^2)$ satisfies the recurrence relation

$$
\mathbb{E}(E_{n+1}^2) = \frac{n+1}{4n+2} \left[ \mathbb{E}(E_n^2) - 4\mathbb{E}(M_n) + 4n + 2 \right].
$$

Now, if $\lim_{n \to \infty} \mathbb{E}(E_n^2) = c$ exists it satisfies the following equation

$$
c = \lim_{n \to \infty} \frac{n+1}{4n+2} \left[ c + 4 \lim_{n \to \infty} \mathbb{E}(M_n) - n \right] + 2.
$$

Since $\lim_{n \to \infty} \mathbb{E}(M_n) - n = -\frac{1}{3}$ from (11) we find $c = \frac{10}{9}$ and therefore

$$
\text{var}(E_n) \sim \frac{10}{9}.
$$

References


Author’s address: E. Stoimenova, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev str., bl. 8, 1113 Sofia, Bulgaria, e-mail: jeni@math.bas.bg.