Theorem 1.1. Let the weight of a word v of length k be $weight(v) := \prod_{i=1}^{k} x[v_i]$. Then the multivariate generating function for words avoiding the consecutive pattern $12...r$ is $\frac{1}{(1-e_1+e_r-e_{r+1}+e_{2r}-e_{2r+1}+...)}$.

Proof. We use the cluster method as described in [NZ]. Let M be the set of marked words on the alphabet $\{1,..,n\}$. And let the weight of a marked word $w := w_1w_2...w_k$ be $weight(w) := (-1)^{|S|} \cdot \prod_{i=1}^{k} x[w_i]$ where S is the set of marks in w. Recall that $weight(M) = weight(M) \cdot (x_1 + x_2 + ... + x_n) + weight(M) \cdot weight(C) + 1$ where C is the set of all possible clusters. The multivariate generating function for words avoiding the consecutive pattern 12...r is equal to $weight(M) = \frac{1}{1-e_1-weight(C)}$. Since the pattern to be avoided is $12..r$, the clusters can only be of the form $(a_1,..,a_j;[1,r],..)$ where $1 \le a_1 < a_2 < ... < a_j \le n$. So $weight(C)$ is a summations of multivariate monomials on $x_1, x_2, .., x_n$ where the exponent of each variable x_i is zero or one.

Now, for any fixed monomial in $weight(C)$, it can come from many different clusters. The number of clusters it comes from and the coefficient of the monomial are uniquely determined by the number of variables in the monomial. For example, for $r = 3$, the monomial $x_1x_3x_5x_6x_7$ can come from the cluster $(13567; [1,3],[2,4],[3,5])$ or $(13567; [1,3],[3,5])$. The first cluster contributes weight $(-1)^3x_1x_3x_5x_6x_7$ whereas the second cluster contributes weight $(-1)^2x_1x_3x_5x_6x_7$. So when summing up, they cancel each other out and there is no monomial $x_1x_3x_5x_6x_7$ in $weight(C)$. So is the case with any other monomial of five variables. Therefore, let us focus on the monomial $x_1x_2x_3..x_k$ and figure out its coefficient. It is clear that for $k < r$ the coefficient of $x_1x_2x_3..x_k$ is 0 because $12...k$ cannot be a cluster (it does not have a mark). And when $k = r$, we have $\text{coeff}(x_1x_2...x_k) = -1$ since there is only one mark. So let us move on to the case when $k > r$.

Claim 1: For $k > r$, $\text{coeff}(x_1x_2...x_k) = -\text{coeff}(x_2x_3...x_k) - \text{coeff}(x_3x_4...x_k) - ... - \text{coeff}(x_rx_{r+1}...x_k)$. (Therefore $\text{coeff}(x_1x_2...x_k) = -\text{coeff}(x_1x_2...x_{k-1}) - \text{coeff}(x_1x_2...x_{k-2}) - ... - \text{coeff}(x_1x_2...x_{k-r+1})$.)

This is because there are $(r-1)$ "categories" of clusters where $x_1x_2...x_k$ can come from, depending on where the second mark in the cluster is. For example, if the clusters are of the form $(1..k; [1,r],[3,r+2],...)$, then the contribution from this "category" of clusters will be $(-1)\cdot$ $\text{coeff}(x_3x_4...x_k)$, with the -1 coming from the first mark $[1,r]$. And if the clusters are of the form $(1..k; [1,r],[r,2r-1],...)$, then the contribution from this "category" will be $(-1)\cdot$ $\text{coeff}(x_rx_{r+1}...x_k)$. Note that if $k < 2r-1$, there cannot be as many as $(r-1)$ "categories" because the second mark can only be at less than $(r-1)$ places. However, in this case, we can "fake" that there are $(r-1)$ places for the second mark because for $k < r$ the coefficient of $x_1x_2x_3..x_k$ is 0. So the above formula still holds. For example, for the clusters associated with the word 123456, and $r = 4$, the first mark has to be 1234, the second mark can only be 2345 or 3456.

But we can "fake" the second mark can also start with 4 and be 456. So coeff($x_1x_2x_3x_4x_5x_6$) $=$ $-$coeff($x_2x_3x_4x_5x_6$)$-$coeff($x_3x_4x_5x_6$)$-$coeff($x_4x_5x_6$)$=$ $-$coeff($x_2x_3x_4x_5x_6$)$-$coeff($x_3x_4x_5x_6$).

We should note here that this idea is not brand new. Readers can find similar ideas (writing $weight(C)$ as a summation of $weight(C[v])$ and $weight(C[v])$ as a summation of $weight(C[u])$ where u is a mark that overlaps with v) on page 7-9 in the the paper [NZ].

So we have: coeff($x_1x_2...x_r$) $= -1$; coeff($x_1x_2...x_{r+1}$) $= (-1)\cdot(-1) = 1$; coeff($x_1x_2...x_{r+2}$) $=$ $-$coeff($x_2x_3...x_{r+2}$) $-$ coeff($x_3x_4...x_{r+2}$) $=$ $-$coeff($x_1x_2...x_{r+1}$) $-$ coeff($x_1x_2...x_r$) $= 0$. Continue this process it is easy to see that $x_1x_2...x_{mr}$ $(m \geq 1)$ has coefficient -1 (so is any other symmetric monomial of mr variables) and $x_1x_2...x_{mr+1}$ has coefficient 1 (so is any other symmetric monomial of mr+1 variables). And the monomials with other number of variables all have coefficient 0. From this argument and summing over all clusters, we conclude $weight(C) = -e_r + e_{r+1} - e_{2r} + e_{2r+1} + ...$ and therefore $weight(M) = \frac{1}{(1-e_1+e_r-e_{r+1}+e_{2r}-e_{2r+1}+...)}$.

Theorem 1.2. Let $F(x_1,\ldots,x_n;t)$ be the multivariate generating function in in $x_1,\ldots,x_n$, whose coefficient of ${x_1}^{m_1}\cdots{x_n}^{m_n}$ is the (one-variable) generating function, in t, such that, for any a, its coefficient of t^a is the number of words m_1 1-s, $\ldots$, m_n n with exactly a occurrences of the pattern $1\ldots r$.
Then

$$F(x_1,\ldots,x_n;t)\frac{1}{1-e_1-(a_re_r+a_{r+1}e_{r+1}+...a_ne_n)} \quad ,$$

where $a_r = t-1$; $a_{r+1} = (t-1)a_r = (t-1)^2$; $a_{r+2} = (t-1)(a_r+a_{r+1})$; ... $a_{2r-1} = (t-1)(a_r+a_{r+1}...+a_{2r-2})$; $a_{2r} = (t-1)(a_{r+1}+a_{r+2}...+a_{2r-1})$; ... $a_n = (t-1)(a_{n-r+1}+...+a_{n-1})$ (In other words, a_i is $(t-1)$ multiplied by the summation of previous $(r-1)$ a_j's with $a_j = 0$ if $j < r$ and $a_r = t-1$.)

Proof. This is a direct generalization from Theorem 1.1. Using the idea as described in page 11 of [NZ], we still let the set of marked words on $\{1,2,...,n\}$ be M. However, this time we let the weight of a marked word w of length k be $weight(w) := (t-1)^{|S|}\cdot\prod_{i=1}^{k} x[w_i]$ where S is the set of marks in w. We still have $weight(M) = weight(M)\cdot(x_1+x_2+...+x_n) + weight(M)\cdot weight(C)+1$ and the multivariate generating function for words having t consecutive patterns $12...r$ is $weight(M) = \frac{1}{1-e_1-weight(C)}$.

The procedure to calculate $weight(C)$ also directly generalizes from the one in Theorem 1.1. We simply replace (-1) by $(t-1)$ in various places, because the only difference is now we assign a different weight to a marked word. For example, we have coeff($x_1x_2...x_r$) $= t-1$;

coeff$(x_1x_2...x_{r+1}) = (t-1)(t-1) = (t-1)^2$; coeff$(x_1x_2...x_{r+2}) = (t-1)($coeff$(x_2x_3...x_{r+2}) +$ coeff$(x_3x_4...x_{r+2})) = (t-1)((t-1)+(t-1)^2)$.

In general, we generalize Claim 1 to the following:

Claim 2: For $k > r$, coeff$(x_1x_2...x_k) = (t-1)$ (coeff$(x_2x_3...x_k)$+ coeff$(x_3x_4...x_k)$ + ...+ coeff$(x_rx_{r+1}...x_k)$). (Therefore coeff$(x_1x_2...x_k) = (t-1)$ (coeff$(x_1x_2...x_{k-1})$+ coeff$(x_1x_2...x_{k-2})$+ ...+ coeff$(x_1x_2...x_{k-r+1})$.)

The proof of Claim 2 directly generalizes directly from the proof of Claim 1 because the only difference is in the weight of a marked word. Now one mark contributes a factor of $(t-1)$ instead of -1 to the weight of a marked word.

Theorem 1.2. then follows from this claim straightforwardly.

Theorem 2. The multi-variate cluster generating function for words on the alphabet $\{1,2,...,n\}$ that avoid $\{[1,2,...,n],[2,3,...,n,1],...,[n,1,2,...,n-1]\}$ is $\frac{\prod_{i=1}^{n} x_i(-n+\sum_{j=1}^{n} x_j)}{1-\prod_{k=1}^{n} x_k}$.

Proof. We rewrite the generating function as $x_1x_2...x_n(-n+x_1+x_2+...+x_n)(1+x_1x_2...x_n+x_1^2x_2^2...x_n^2+...)$. And we will use the case $n=3$ to illustrate the idea.

First of all, no matter how long the cluster is, it will always be of the form $\{a_1a_2a_3a_1a_2a_3a_1...\}$ where $a_1a_2a_3 \in \{[1,2,3],[2,3,1],[3,1,2]\}$. If we start with 123, for example, in our cluster, and say, our cluster has 7 letters in total, then it has to be 1231231. Why is it so? That is because, any two adjacent letters must belong to one mark (by the definition of cluster). Therefore the letters proceed in the "clockwise" fashion. That is, after 3 there must be a 1, after 1 there must be a 2 and after 2 there must be a 3. Having observed what the clusters look like, now let us figure out what are the coefficients for the clusters.

Say our cluster is of length $3m$ (m is a positive integer), and it starts with 123. From the argument above, we know it is 123123...123 (123 repeated m times). We use induction to show that the coefficient of corresponding term in the generating function $x_1^m x_2^m x_3^m$ is -1. If $m=1$, obviously we just have one mark, so the coefficient of $x_1x_2x_3$ is -1. Now look at our cluster 123123...123 (123 repeated m times). There could be three potential marks overlapping the first 123. They are 123, 231 and 312. Of the marks 231, 312, we have to choose at least one of them (we cannot choose neither of them because then we will not have a cluster). If we just choose one of them to be a mark, the contribution to the coefficient of $x_1^m x_2^m x_3^m$ will be $(+1)$ multiplied by the coefficient of $x_1^{m-1}x_2^{m-1}x_3^{m-1}$. If we choose both of them as marks, the contribution will be (-1) multiplied by the coefficient of $x_1^{m-1}x_2^{m-1}x_3^{m-1}$. So in total, the contribution will be $(\binom{2}{1} - \binom{2}{2})$ multiplied by the coefficient of $x_1^{m-1}x_2^{m-1}x_3^{m-1}$ (which is -1 by induction hypothesis). So we get -1 as the contribution from 123123...123 to coefficient of $x_1^m x_2^m x_3^m$. Since the other length 3m clusters are 231231...231 and 312312...312 (each contributing -1 to the coefficient), we get -3 as the coefficient of $x_1^m x_2^m x_3^m$. (This works the same way for the general case, because of the fact that $\binom{k}{1} - \binom{k}{2} + \binom{k}{3} + .. + (-1)^{k+1}\binom{k}{k} = 1$ for any positive integer k).

If our cluster is of length $3m+1$ and say it starts with 123, we know it is 123123...1231 (123 repeated m times). We can again use induction to show that the coefficient of $x_1^{m+1}x_2^m...x_n^m$ is $+1$. The base case 1231 now contributes $+1$ to the coefficient of $x_1^2x_2x_3$. Also note that the only cluster that $x_1^{m+1}x_2^m...x_n^m$ comes from is 123123...1231, so we have one case instead of 3. By similar induction argument as above we see that the coefficient of $x_1^{m+1}x_2^m x_3^m$ is $+1$. And so is the coefficient of $x_1^m x_2^{m+1} x_3^m$ and $x_1^m x_2^m x_3^{m+1}$.

Now another possibility is our cluster is of length $3m+2$. Again say it starts with 123, then it must be 123123...12312 (123 repeated m times). Why do we not see any $x_1^{m+2}x_2^m...x_n^m$ term in the generating function? This is because the base case 12312 contributes 0 to the coefficient $x_1^2x_2^2x_3$ (which can only come from 12312). In general, the contribution from $123..n12..r$ $(1 < r < n)$ to the coefficient $x_1^2x_2^2...x_r^2x_{r+1}...x_n$ (which can only come from $123..n12..r$) will be 0. This is because we can choose any number of marks from $\{23..n1, 34..n12, ..., r(r+1)...(r-2)(r-1)\}$ and $-\binom{k}{0} + \binom{k}{1} - \binom{k}{2} + \binom{k}{3} + .. + (-1)^{k+1}\binom{k}{k} = 0$ for any positive integer k).

Gathering the cases above, we get the generating function stated in Theorem 2.

Theorem 3. The multi-variate cluster generating function for words on alphabet $\{1, 2, ..., n\}$ avoiding any pattern in $\{[2,3,1], [3,1,2], [2,1,3], [1,3,2]\}$ has the same number of terms (unlike pattern [1,2,3] and [3,2,1], the generating functions for these patterns are not symmetric).

Proof. This is because of the simple fact that any two patterns of $\{[2,3,1], [3,1,2], [2,1,3], [1,3,2]\}$ can be transformed to one another by the operation of reversion or taking complement. And we can apply the same operation to the cluster we are looking at to establish a bijection. For example, 231 is the complement of 213 (meaning 2 goes to 2, 1 goes to 3, 3 goes to 1). If we are looking at clusters on $\{1, 2, 3, 4\}$ related to pattern 213 and 231 respectively, the cluster 21324 translates to cluster 34231. In other words there is a bijection from the clusters related to the pattern 213 and the clusters related to the pattern 231.