K-S Test

Aytijhya Saha , Samahriti Mukherjee

16 August 2021

Show that to study K-S Statistic which requires only the min/max value, under the assumption that m=n, null hypothesis and underlying distribution is continuous, it is enough to study maximum of a SSRW.

As the underlying distribution is continuous, therefore all the sample points, $\{X_1, X_2, .., X_n, Y_1, Y_2, .., Y_n\}$ are distinct with probability 1.
Let,$\{X_1, X_2, .., X_n, Y_1, Y_2, .., Y_n\} = A$
Let,$\{X_1, X_2, .., X_n\} = B$ and $\{Y_1, Y_2, .., Y_n\} = C$
So, we arrange the elements of A in increasing order
Let $Z_1 < Z_2 < .... < Z_{2n}$ be the elements of A. For an interpretation in terms of paths,we write $\epsilon_p = +1$ or -1 according as $Z_p \in A$ or $Z_p \in B$. According to null hypothesis, elements of A are coming from the same distribution function.
So $P[Z_p = X_i] = P[Z_p = Y_j] \forall i, j \in [1, n], p \in [1, 2n]$.
So,$P[Z_p \in B] = P[Z_p \in C] = \frac{1}{2}$. Hence, the random walk is simple symmetric.
Let, $t \in [z_k, z_{k+1})$, for some k.

Now when $|\hat{F_n}(t) - \hat{G_n}(t)| > c$, then $|s_k| = n|\hat{F_n}(t) - \hat{G_n}(t)| > cn$

When given that, $|s_k| > cn$ for some k,take $t = z_k$, then, $|\hat{F_n}(t) - \hat{G_n}(t)| > c$.(proved)

Let, our test statistic, $sup_{x\in\mathbb{R}}|F_n\hat{(}x) - G_n\hat{(}x)| \leq M$, then
$F_n\hat{(}x) - G_n\hat{(}x) \leq M \forall x$
$\iff$ $|s_k| \leq Mn$ for all $k \in \{1, 2, .., n\}$, as if $|s_k| > Mn$ for some k, then $|F_n\hat{(}x) - G_n\hat{(}x)| > Mn$ by our claim which we proved earlier, but M is the supremum of $F_n\hat{(}x) - G_n\hat{(}x)$
$\iff$ $sup\{|s_k| : k = 1, .., n\} \leq nM$.
So,it is enough to study maximum of a SSRW.