Proof of Theorem 1.
Take some time $T_3 > T_2$, and assume that the window has not been cut since $T_0$. We have three segments in the log represented in the window:

- before $T_0..T_1$, with $P$ having stable mass $\mu_1$.
- from $T_2..T_3$ onwards, with $P$ having stable mass $\mu_2$
- between $T_1..T_2$, where we have no information about how the mass (it may be changing wildly). But let’s say the average mass in this segment is $\mu$.

Say to simplify cases that $\mu_2 > \mu_1$. Now one of two happens: either $\mu \leq (\mu_2 + \mu_1)/2$ or $\mu \geq (\mu_2 + \mu_1)/2$. Say we are in the first case (the second case is symmetrical), then $\mu_2 - \mu \geq (\mu_2 - \mu_1)/2$.

Now, we can split ADWIN’s window at time $T_3$ into two parts: a sub-window of length $T_2 - T_0$ and another one of length $T_2 - T_2$. The expected average in the second segment is $\mu_1$, and by the above the expected average in the first segment is in somewhere between $\mu_1$ and $\mu$, which by the above is less than $\mu_2 - (\mu_2 - \mu_1)/2$. So the difference in their averages is at least $(\mu_2 - \mu_1)/2$ in expectation. By theorem 3.1 in BifetG07, part (b), ADWIN with confidence parameter $\delta$ will detect change with probability at least $1 - \delta$ if

$$
\frac{(\mu_2 - \mu_1)^2}{16 \ln(4(T_3 - T_0)/\delta)} \geq \frac{1}{T_2 - T_0} + \frac{1}{T_3 - T_2}.
$$

If $T_1 - T_0(<T_2 - T_0)$ is sufficiently large, the first term on the right hand side can be ignored or subsumed by the second, and the above is true if

$$
T_3 \geq T_2 + \frac{16 \ln(4(T_3 - T_0)/\delta)}{(\mu_2 - \mu_1)^2}
$$

and approximating $T_3$ by $T_2$ inside the logarithm at the expense of a somewhat larger constant $c$, this is true for

$$
T_3 \geq T_2 + \frac{c \ln(4(T_2 - T_0)/\delta)}{(\mu_2 - \mu_1)^2}
$$

which is the statement of the theorem, for fixed parameter $\delta$. 

1