

# LARGE DEVIATION PRINCIPLES FOR WORDS DRAWN FROM CORRELATED LETTER SEQUENCES

F. DEN HOLLANDER AND J. POISAT

**ABSTRACT.** When an i.i.d. sequence of letters is cut into words according to i.i.d. renewal times, an i.i.d. sequence of words is obtained. In the *annealed* LDP (large deviation principle) for the empirical process of words, the rate function is the specific relative entropy of the observed law of words w.r.t. the reference law of words. In Birkner, Greven and den Hollander [3] the *quenched* LDP (= conditional on a typical letter sequence) was derived for the case where the renewal times have an *algebraic* tail. The rate function turned out to be a sum of two terms, one being the annealed rate function, the other being proportional to the specific relative entropy of the observed law of letters w.r.t. the reference law of letters, obtained by concatenating the words and randomising the location of the origin. The proportionality constant equals the tail exponent of the renewal process.

The purpose of the present paper is to extend both LDP's to letter sequences that are not i.i.d. It is shown that both LDP's carry over when the letter sequence satisfies a mixing condition called *summable variation*. The rate functions are again given by specific relative entropies w.r.t. the reference law of words, respectively, letters. But since neither of these reference laws is i.i.d., several approximation arguments are needed to obtain the extension.

*MSC* 2010: 60F10, 60G10.

*Key words:* Letters and words, renewal times, empirical process, annealed vs. quenched large deviation principle, rate function, specific relative entropy, mixing.

*Acknowledgment:* FdH and JP were supported by ERC Advanced Grant 267356 VARIS.