Statistical properties of entropy production derived from fluctuation theorems

Neri Merhav[(1)] **and Yariv Kafri**[(2)]

(1) Department of Electrical Engineering, Technion, Haifa 32000, Israel. (2) Department of Physics, Technion, Haifa 32000, Israel

Abstract. Several implications of well–known fluctuation theorems, on the statistical properties of the entropy production, are studied using various approaches. We begin by deriving a tight lower bound on the variance of the entropy production for a given mean of this random variable. It is shown that the Evans–Searles fluctuation theorem alone imposes a significant lower bound on the variance only when the mean entropy production is very small. It is then nonetheless demonstrated that upon incorporating additional information concerning the entropy production, this lower bound can be significantly improved, so as to capture extensivity properties. Another important aspect of the fluctuation properties of the entropy production is the relationship between the mean and the variance, on the one hand, and the probability of the event where the entropy production is negative, on the other hand. Accordingly, we derive upper and lower bounds on this probability in terms of the mean and the variance. These bounds are tighter than previous bounds that can be found in the literature. Moreover, they are tight in the sense that there exist probability distributions, satisfying the Evans–Searles fluctuation theorem, that achieve them with equality. Finally, we present a general method for generating a wide class of inequalities that must be satisfied by the entropy production. We use this method to derive several new inequalities which go beyond the standard derivation of the second law.

1. Introduction

It has been recently realized that time reversal symmetry implies that some exact results, termed *fluctuation theorems*, hold for systems which are driven arbitrarily far from thermal equilibrium [1].

As an example, which will be of particular interest in this paper, consider an isolated system in a certain equilibrium state, which is driven to another state by varying an external control parameter λ over a time interval $0 \leq t \leq T$, according to a certain protocol $\{\lambda(t),\ 0 \leq t \leq T\}$. Following the application of the protocol, a difference σ is developed between the entropy at the final state and the entropy at the initial state of the system. This difference is also called the *entropy production.* One can now imagine the same system being driven using a time–reversed protocol $\tilde{\lambda}(t) = \lambda(T-t)$. When the time–reversed protocol obeys $\tilde{\lambda}(t) = \lambda(t)$, it can be shown that time reversal symmetry implies the Evans–Searles fluctuation theorem [2, 3]

$$\frac{p(\sigma)}{p(-\sigma)} = e^{\sigma} , \tag{1}$$

where $p(\sigma)$ is the probability density function of the entropy production σ. This immediately yields

$$\left\langle e^{-\sigma} \right\rangle = 1, \tag{2}$$

where as usual, we assume that the entropy change in the system is much smaller than its total entropy. As is well known, the second law of thermodynamics, $\langle \sigma \rangle \geq 0$, is easily obtained by applying Jensen's inequality, $\langle e^{-\sigma} \rangle \geq e^{-\langle \sigma \rangle}$, to the left–hand side of eq. (2). Other, directly related examples of fluctuation theorems include the Jarzynski equality, the Crooks theorem, and the Gallavotti–Cohen relation [5, 6, 7, 9, 10, 11, 12, 13].

Beyond their elegance and their pure academic value, some of the fluctuation theorems have been suggested as possible tools for measuring *equilibrium* quantities from averages over repeated experiments of a *non–equilibrium* process [1], which is a considerably interesting idea. For instance, the Jarzynski equality paves the way to measure free–energy differences from non–equilibrium measurements of work performed on (or by) the system. However, as is evident from eq. (2), events with a negative entropy production play an important role, which makes such a measurement impossible for macroscopic systems.

As mentioned above, when Jensen's inequality is applied to eq. (2), the second law is immediately obtained. Moreover, since the exponential function is strictly convex, Jensen's inequality becomes an equality if and only if σ is a degenerate random variable, which takes on the value $\sigma = 0$ with probability one. This corresponds to a perfectly reversible process where no entropy is produced. In any other case, where Jensen's inequality is a strict one, σ becomes a non–degenerate random variable, and hence it fluctuates about its mean, which then must be strictly positive. Thus, eq. (2) already tells us an important qualitative fact: Irreversibility, strictly positive mean entropy production, and fluctuations (e.g., in terms of positive variance) appear always together. However, eq. (2) does not immediately tell us much in the quantitative level. In particular, when applying Jensen's inequality, it is felt that a great deal of valuable information, concerning the statistics of the entropy production, is lost.

It is the purpose of this paper to explore some quantitative restrictions on the probability distribution of the entropy production that arise from the fluctuation theorems. Throughout most of the paper, we consider systems which obey the Evans–Searles fluctuation theorem, i.e., systems that satisfy eq. (1). Some restrictions are rather obvious. Consider for example the mean entropy production

$$m = \langle\sigma\rangle \tag{3}$$

and the variance

$$v = \langle\sigma^2\rangle - \langle\sigma\rangle^2. \tag{4}$$

In the small mean entropy production limit, one can expand eq. (2) to leading order and obtain the simple relation $v = 2m$. The same restriction holds, of course, also in the Gaussian case (even without taking the small mean entropy limit), as only the first two cumulants are non–zero in this case.

In this paper, we derive the most stringent restrictions, that can possibly be posed by the fluctuation theorems, on the shape of the distribution. In particular, we first derive a lower bound on the variance v of the entropy production σ as a function of the mean entropy production m. The bound we derive is tight in the sense that there is a probability distribution of the entropy production, which achieves the bound with equality. It turns out, from this bound, that the Evans–Searles fluctuation theorem alone imposes a meaningful lower bound on the variance v only when the mean entropy production m is very small. It is nonetheless demonstrated that upon incorporating additional statistical (and/or physical) information concerning the entropy production, this lower bound can be significantly improved, so as to capture extensivity properties, which mean that in the thermodynamic limit, it is plausible that both m and v should scale linearly with time and with the system size, and so, they should be proportional to each other.

Next, we derive upper and lower bounds on the probability of negative entropy production, $\Pr\{\sigma \leq 0\}$, as a function of the mean entropy production m and the variance v. To the best of our knowledge, these bounds are better than related bounds, previously derived using Jensen's inequality [14]. They are again tight in the sense that there exist probability distributions, satisfying the Evans–Searles fluctuation theorem, which achieve them with equality. The interesting fact, in this context, is that we are actually obtaining a bound on the large deviations rate function of the probability of the rare event $\{\sigma \leq 0\}$. As will be shown later, if σ is an extensive random variable, then this probability can decay no faster than e^{-m}, which is exponential in the system size (or time). The upper and lower bounds are also *universal* (in the probability distribution of σ) and the important fact is that they provide non–trivial assessments on $\Pr\{\sigma \leq 0\}$, a quantity which is not easily measurable by experiments, in terms of m and v, which are measurable in principle.

Our last result is about a general analysis tool to be applied to eq. (2) in order to derive a wide class of inequalities that involve the entropy production. These inequalities are, in general, more powerful than the standard Jensen inequality, and some of them lead to certain variations of the second law, $\langle\sigma\rangle \geq 0$. For example, among other things, we prove that the probability of the event $\{\sigma \leq \alpha\}$, for any deterministic parameter α, is bounded from above according to

$$\Pr\{\sigma \leq \alpha\} \leq \exp\left\{\langle\sigma\rangle|_{\sigma\leq\alpha}\right\}, \tag{5}$$

where the average is conditional on $\sigma \leq \alpha$. Thus, the second law is obtained as a special case, with $\alpha \to \infty$. Another example of an inequality from this class is

$$\langle\sigma e^{-\sigma}\rangle \leq 0, \tag{6}$$

which tells us that although σ is non–negative on the average, as the second law asserts, the event of negative entropy production still has enough probabilistic weight so as to make the average of σ, weighted by the function $e^{-\sigma}$ (which favors negative values of σ), negative rather than positive.

The remaining part of the paper is structured as follows: In Sec. 2, we derive lower bounds on the variance of the entropy production given the mean. In Sec. 3, lower and upper bounds are derived for the probability of rare events with a negative entropy production. In Sec. 4, we present the method for deriving inequalities as mentioned above. Finally, we conclude in Sec. 5.

2. A lower bound on the variance of the entropy production

Consider a system which obeys eq. (1), with $\langle\sigma\rangle = m$, where m is given. First, observe that since $p(-\sigma) = e^{-\sigma}p(\sigma)$, we have

$$\int_{-\infty}^{+\infty} \mathrm{d}\sigma \cdot p(\sigma) = \int_{0}^{+\infty} \mathrm{d}\sigma p(\sigma)(1+e^{-\sigma}) = 1, \tag{7}$$

which means that instead of considering a real valued random variable σ, taking both positive and negative values, and distributed according to $p(\sigma)$, we can consider, equivalently, a positive random variable, distributed according to

$$q(\sigma) = p(\sigma)(1+e^{-\sigma}), \quad \sigma \geq 0. \tag{8}$$

The mean m can then be expressed, in terms of q, according to

$$m = \int_{-\infty}^{+\infty} \mathrm{d}\sigma \cdot \sigma p(\sigma) \tag{9}$$

$$= \int_{0}^{+\infty} \mathrm{d}\sigma \cdot p(\sigma)\sigma(1-e^{-\sigma}) \tag{10}$$

$$= \int_{0}^{+\infty} \mathrm{d}\sigma \cdot q(\sigma)\sigma \cdot \frac{1-e^{-\sigma}}{1+e^{-\sigma}} \tag{11}$$

$$= \int_{0}^{+\infty} \mathrm{d}\sigma \cdot q(\sigma)\sigma \tanh\left(\frac{\sigma}{2}\right) \tag{12}$$

$$= \int_{0}^{+\infty} \mathrm{d}\sigma \cdot q(\sigma) f(\sigma), \tag{13}$$

where we have defined

$$f(\sigma) \equiv \sigma \tanh\left(\frac{\sigma}{2}\right). \tag{14}$$

Similarly, the second moment is

$$\left\langle \sigma^2 \right\rangle = \int_0^{+\infty} \mathrm{d}\sigma \cdot \sigma^2 p(\sigma)(1+e^{-\sigma}) = \int_0^{+\infty} \mathrm{d}\sigma \cdot \sigma^2 q(\sigma). \tag{15}$$

In simple words, we have used the Evans–Searles fluctuation theorem in order to transform a two–sided random variable σ, governed by $p(\sigma)$, into a one–sided random variable ($\sigma \geq 0$), whose probability density function is given by $q(\sigma) = p(\sigma)(1+e^{-\sigma})$ and we henceforth denote expectations under p and under q, by $\langle\cdot\rangle_p$ and $\langle\cdot\rangle_q$, respectively. Thus, the constraint $\langle\sigma\rangle_p = m$, in the domain of the original two–sided random variable, is equivalent to the constraint $\langle f(\sigma)\rangle_q = m$ in the domain of the transformed, one–sided random variable.

Since f is a monotonically strictly increasing function for $\sigma \geq 0$, it is clearly invertible in this range, and we shall denote the inverse function of f by h. I.e., $\mu = f(\sigma)$ if and only if $\sigma = h(\mu)$. We next observe that $h^2(\mu) = [h(\mu)]^2$, which is obviously the inverse function of $f(\sqrt{\sigma})$, is a convex function.‡ Thus, we readily obtain the following lower bound on the second moment in terms of m:

$$\left\langle \sigma^2 \right\rangle_p = \left\langle \sigma^2 \right\rangle_q \tag{16}$$

$$= \left\langle h^2(f(\sigma)) \right\rangle_q \tag{17}$$

$$\geq h^2(\langle f(\sigma)\rangle_q) \tag{18}$$

$$= h^2(m), \tag{19}$$

where the inequality follows from the application of Jensen's inequality to the convex function h^2. Equality is obtained when σ is deterministic under q, i.e.,

$$q(\sigma) \equiv q^*(\sigma) = \delta(\sigma - h(m)), \tag{20}$$

or equivalently,

$$p(\sigma) = p^*(\sigma) \equiv \frac{1}{1+e^{-h(m)}} \cdot \delta(\sigma - h(m)) + \frac{e^{-h(m)}}{1+e^{-h(m)}} \cdot \delta(\sigma + h(m)). \tag{21}$$

Using the above result, we see that the variance of the entropy production can be bounded by

$$v \geq h^2(m) - m^2. \tag{22}$$

This is the central result of this section. The function $h^2(m) - m^2$ is depicted in Fig. 1.

We comment that the above analysis holds, not only for the second moment. It can be generalized straightforwardly from the second moment, $\langle\sigma^2\rangle$, to every higher moment of the form $\left\langle |\sigma|^k \right\rangle$, where k is any real number larger than 2 (i.e., k does not have to be an integer). The more general result is then

$$\left\langle |\sigma|^k \right\rangle \geq h^k(m) \equiv [f^{-1}(m)]^k, \tag{23}$$

‡ This follows from the fact that $f(\sqrt{\sigma})$ is monotonically increasing and concave for $\sigma \geq 0$, as can easily be checked from the derivatives of this function.

where equality is universally achieved by the same density function p^* as before.

Returning to the second moment, it is easy to see that $h(x) \geq x$ (with equality only at $x = 0$), so the lower bound on the variance is positive. For small m, $h^2(m) \approx 2m$, and so, to this order, $v \geq 2m$, in agreement with the simple relation obtained by expanding eq. (2). For large values of m, we have to leading order $h(m) = m(1 + 2e^{-m/2})$ and the bound decays as $4e^{-m/2}$. This behavior is easy to understand since the minimum variance distribution, for a given m, is achieved by a pair of delta functions, as seen above in eq. (21). For large values of m, the contribution of the Dirac function at the negative value of σ becomes essentially irrelevant to the variance, due to the exponential weighting of the Evans–Searles theorem. As advertised in the Introduction, in this context, the fluctuation theorems provide useful information on the variance of the entropy production distribution only for relatively small values of the mean entropy production.

The reason for this behavior is simple: The Evans–Searles fluctuation theorem merely relates the probability density p at negative values of σ to those at the corresponding positive values, but as m grows without bound (the thermodynamic limit), most of the probability mass goes for positive values of σ anyway, and so, the information concerning negative values becomes essentially irrelevant. Therefore, it is understood that the Evans–Searles theorem *alone* cannot possibly give useful information about the thermodynamic limit, and this is not because of a possible weakness in the derivation of the lower bound (which is tight, in the sense of being achieved by p^*). This means that in order to obtain more meaningful bounds, one must incorporate additional information.

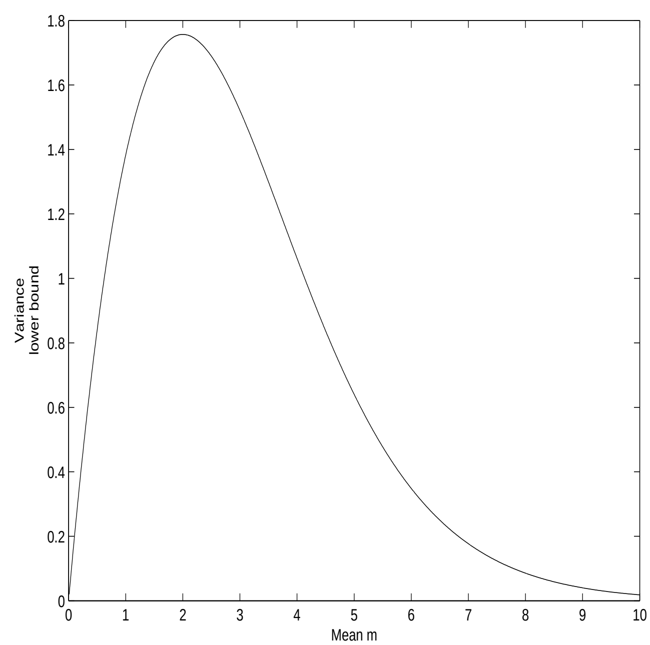

Figure 1. A plot of the lower bound on the variance, $h^2(m) - m^2$, as a function of m.

To demonstrate this fact, suppose that the additional information we have is given by the extra constraint

$$\Pr\{|\sigma| \le r\} = b, \tag{24}$$

where $r > 0$ and $b \in [0, 1]$ are given§ and where we will assume that $f(r) < m$. How does the lower bound on the variance v change in the presence of this additional constraint?

Once again, we refer to the one–sided distribution q. The two constraints now read $\langle f(\sigma)\rangle_q = m$ and $q\{\sigma \le r\} = b$. Next, we note that every probability measure q, satisfying the two constraints on the positive reals, can be represented as a mixture of two probability density functions, q_1 and q_2, as follows:

$$q(\sigma) = bq_1(\sigma) + (1 - b)q_2(\sigma), \tag{25}$$

where the support of q_1 is (a subset of) the interval $[0, r]$ and the the support of q_2 is (a subset of) $(r, \infty]$. We next denote expectations with respect to q_1 and q_2 by $\langle\cdot\rangle_1$ and $\langle\cdot\rangle_2$, respectively. Let us define $m_1 = \langle f(\sigma)\rangle_1$ and $m_2 = \langle f(\sigma)\rangle_2$, where, of course, $bm_1 + (1 - b)m_2 = m$, $m_1 \le f(r)$. Now, from the same considerations as in the proof of the first bound derived above, we have

$$\left\langle \sigma^2 \right\rangle_1 \ge h^2(m_1) \tag{26}$$

and

$$\left\langle \sigma^2 \right\rangle_2 \ge h^2(m_2). \tag{27}$$

This implies

$$\begin{aligned}
\left\langle \sigma^2 \right\rangle &= b\left\langle \sigma^2 \right\rangle_1 + (1 - b)\left\langle \sigma^2 \right\rangle_2 \\
&\ge bh^2(m_1) + (1 - b)h^2(m_2) \\
&= bh^2(m_1) + (1 - b)h^2\left(\frac{m - bm_1}{1 - b}\right) \\
&\ge bh^2(f(r)) + (1 - b)h^2\left(\frac{m - bf(r)}{1 - b}\right) \\
&= br^2 + (1 - b)h^2\left(\frac{m - bf(r)}{1 - b}\right),
\end{aligned} \tag{28}$$

where the second inequality follows from the limitation $m_1 \le f(r)$ and the convexity of h^2, which implies that the minimum of $bh^2(m_1) + (1 - b)h^2\left(\frac{m-bm_1}{1-b}\right)$, subject to the constraint $m_1 \le f(r)$, is achieved for $m_1 = f(r)$ (provided that $f(r) < m$, as assumed). The above lower bound is achieved as both inequalities in the last chain become equalities, and this is the case for

$$q(\sigma) = b\delta(\sigma - r) + (1 - b)\delta\left(\sigma - h\left(\frac{m - bf(r)}{1 - b}\right)\right), \tag{29}$$

which means two pairs of Dirac delta functions in the domain of the original, one–sided random variable, governed by p.

§ In practice, this additional information may be a result of an experimental measurement.

In view of the foregoing discussion on the first lower bound, it is now interesting to examine what happens in the limit of large mean entropy production m. In this regime, it makes sense to let r increase with m, so as to keep b approximately constant and thus to avoid a situation where the event $|\sigma| \le r$ becomes a rare one. In the lower bound, when both r and $m - bf(r)$ are large, $f(r) \approx r$ and

$$h\left(\frac{m - bf(r)}{1-b}\right) \approx \frac{m - bf(r)}{1-b} \approx \frac{m - br}{1-b} \tag{30}$$

and so,

$$\begin{aligned} \left\langle \sigma^2 \right\rangle &\ge br^2 + (1-b)\left(\frac{m - br}{1-b}\right)^2 \\ &= \frac{br^2 + m^2 - 2mrb}{1-b} \qquad (31) \\ &= m^2 + \frac{b(r-m)^2}{1-b}, \qquad (32) \end{aligned}$$

which means that the asymptotic lower bound on the variance is

$$v \ge \frac{b(r-m)^2}{1-b}. \tag{33}$$

This is especially suitable in situations where central limit theorem arguments hold, and then, in the vicinity of the peak, the distribution can be approximated by a Gaussian with a mean proportional to the variance (both being extensive variables). Then b remains approximately constant if the dependence of r upon m is chosen to be $r = r(m) \equiv m - c\sqrt{m}$, where c is some arbitrary positive constant. In this case, the lower bound (33) becomes

$$v \ge \frac{bc^2 m}{1-b}, \tag{34}$$

which is indeed extensive (proportional to m), as may be expected. This is very different from the earlier lower bound, which vanishes as m tends to infinity.

3. Bounds on the probability of negative entropy production

Eq. (1) embeds a certain symmetry property of the entropy production about the origin. In this section, we employ this symmetry to derive upper and lower bounds on the probability of the (rare) event that the entropy decreases. This serves as another measure of fluctuations in the entropy production. Namely, we derive bounds on $\Pr\{\sigma \le 0\} = \int_{-\infty}^{0} d\sigma p(\sigma)$, which depend merely on the mean m and the variance v.‖

For the lower bound, we use the simple inequality

$$e^{-x} \ge e^{-\alpha} - e^{-\alpha}(x - \alpha), \tag{35}$$

‖ More precisely, our lower bound will depend merely on m, whereas the upper bound will depend on both m and v.

which follows from the simple fact that the exponential function is lower bounded by the affine function tangential to it at any point $(\alpha, e^{-\alpha})$ on the curve, where α is an arbitrary real parameter, whose value is left for our choice. Using this inequality, we now have

$$\begin{aligned}
\Pr\{\sigma \leq 0\} &= \int_{-\infty}^{0} \mathrm{d}\sigma \cdot p(\sigma) \\
&= \int_{0}^{\infty} \mathrm{d}\sigma e^{-\sigma} \cdot p(\sigma) && (36)\\
&\geq \int_{0}^{\infty} \mathrm{d}\sigma [e^{-\alpha} - e^{-\alpha}(\sigma - \alpha)] \cdot p(\sigma) && (37)\\
&= (1+\alpha)e^{-\alpha}\Pr\{\sigma > 0\} - e^{-\alpha}\int_{0}^{\infty} \mathrm{d}\sigma \cdot \sigma p(\sigma) && (38)\\
&= (1+\alpha)e^{-\alpha}[1 - \Pr\{\sigma \leq 0\}] - e^{-\alpha}\int_{0}^{\infty} \mathrm{d}\sigma \cdot \sigma p(\sigma), && (39)
\end{aligned}$$

and so

$$\Pr\{\sigma \leq 0\} \geq \frac{(1+\alpha-m_+)e^{-\alpha}}{1+(1+\alpha)e^{-\alpha}} = \frac{1+\alpha-m_+}{e^{\alpha}+\alpha+1}, \qquad (40)$$

where we have denoted $m_+ = \int_0^\infty \mathrm{d}\sigma \cdot \sigma p(\sigma)$. Since this holds for any real α, then

$$\Pr\{\sigma \leq 0\} \geq \sup_{\alpha} \frac{1+\alpha-m_+}{e^{\alpha}+\alpha+1}. \qquad (41)$$

The maximum is achieved for $\alpha = \psi(m_+)$, where $\psi(\cdot)$ is the inverse of the function $\phi(\alpha) = \alpha/(1+e^{-\alpha})$. This yields

$$\Pr\{\sigma \leq 0\} \geq \frac{1+\psi(m_+)-m_+}{e^{\psi(m_+)}+\psi(m_+)+1}. \qquad (42)$$

In the the thermodynamic limit, $m_+ \approx m$ and $\alpha \approx m_+$, and so,

$$\Pr\{\sigma \leq 0\} \geq \frac{1}{e^{m}+m+1} \sim e^{-m}. \qquad (43)$$

We observe then that in the thermodynamic limit, if m is an extensive variable, the probability of a negative entropy production cannot decay exponentially faster than e^{-m}. While this lower bound is universal (in the sense of being independent of the actual probability distribution of σ), it is nevertheless a tight bound in the sense that it is achieved by a certain distribution that satisfies the Evans–Searles fluctuation theorem (again, given by a pair of Dirac delta functions).

Next we derive an upper bound on the probability of negative entropy production. To this end, we use the following inequality which applies for any $x \geq 0$ and $a \geq 0$:

$$e^{-x} \leq e^{-a} - e^{-a}(x-a) + g(x-a)^2 \qquad (44)$$

where

$$g \equiv \frac{1-(a+1)e^{-a}}{a^2}\ . \qquad (45)$$

This upper bounds the exponential function e^{-x} by a quadratic function, tangential to the exponential function at the point $x = a$, where the coefficient g of the quadratic term is chosen so as to keep the quadratic function above the exponential function for every positive x. Denoting $s_+ = \int_0^\infty \mathrm{d}\sigma \cdot \sigma^2 p(\sigma)$, we have

$$\Pr\{\sigma \le 0\} = \int_0^\infty \mathrm{d}\sigma e^{-\sigma} \cdot p(\sigma) \tag{46}$$

$$\le \int_0^\infty \mathrm{d}\sigma \left[e^{-a} - e^{-a}(\sigma - a) + g(\sigma^2 - 2a\sigma + a^2)\right] \cdot p(\sigma) \tag{47}$$

$$= [(a+1)e^{-a} + a^2 g]\Pr\{\sigma > 0\} - e^{-a}m_+ + g(s_+ - 2am_+) \tag{48}$$

$$= \Pr\{\sigma > 0\} - e^{-a}m_+ + g(s_+ - 2am_+) \tag{49}$$

$$= 1 - \Pr\{\sigma \le 0\} - e^{-a}m_+ + g(s_+ - 2am_+) \tag{50}$$

and so,

$$\Pr\{\sigma \le 0\} \le \frac{1 + gs_+ - 2agm_+ - e^{-a}m_+}{2}. \tag{51}$$

Since this holds for every $a \ge 0$,

$$\Pr\{\sigma \le 0\} \le \inf_{a\ge 0} \frac{1 + gs_+ - 2agm_+ - e^{-a}m_+}{2} \tag{52}$$

$$= \frac{1}{2} + \frac{1}{2}\inf_{a\ge 0}(gs_+ - 2agm_+ - e^{-a}m_+). \tag{53}$$

A minimization over a yields $a = s_+/m_+$ so that,

$$gs_+ - 2agm_+ - e^{-a}m_+ = gs_+ - 2gs_+ - e^{-a}m_+ \tag{54}$$

$$= -(gs_+ + e^{-a}m_+) \tag{55}$$

$$= -(gs_+ + m_+ e^{-s_+/m_+}) \tag{56}$$

where

$$g = \frac{m_+^2}{s_+^2} - \left(\frac{m_+}{s_+} + \frac{m_+^2}{s_+^2}\right) e^{-s_+/m_+}.$$

Thus,

$$gs_+ + m_+ e^{-s_+/m_+} = \frac{m_+^2}{s_+}\left(1 - e^{-s_+/m_+}\right) \tag{57}$$

and so,

$$\Pr\{\sigma \le 0\} \le \frac{1}{2} - \frac{m_+^2}{2s_+} + \frac{m_+^2}{2s_+} e^{-s_+/m_+}. \tag{58}$$

Similarly as the lower bound, the upper bound too is universal and tight in the sense defined above. Here, in the thermodynamic limit, $m_+ \to m$ and $s_+ \to s \equiv v + m^2$, which yields

$$\Pr\{\sigma \le 0\} \le \frac{1}{2}\left(1 - \frac{m^2}{s}\right) + \frac{m^2}{2s} e^{-s/m} \tag{59}$$

$$= \frac{v}{2s} + \frac{m^2}{2s} e^{-s/m}. \tag{60}$$

Using the bound obtained in the previous section (see eq. (22) and the discussion following it), one may wonder what is the range of values that this bound can possibly take for a given values of m and v. One can check that the minimum of the bound, in the large m limit, is given when the variance is minimum, i.e., $v = 4e^{-m/2}$, so that to leading order, it is $2e^{-m/2}/m^2$. Its maximum value is achieved when $v \gg m^2$, and it behaves like $v/(2s) = 1/(2m^2/v + 2)$. For comparison, we note that the well–known Chebychev inequality (that also bounds the probability of interest in terms of m and v) yields

$$\begin{aligned} \Pr\{\sigma \leq 0\} &= \Pr\{m - \sigma \geq m\} && (61)\\ &\leq \frac{\langle (m-\sigma)^2 \rangle}{m^2} && (62)\\ &= \frac{v}{m^2}, && (63) \end{aligned}$$

which is a weaker upper bound, that does not make use of the fluctuation theorem.

4. Inequalities based on Jensen's inequality with a change of measure

As mentioned already in the Introduction, it is common practice to go from eq. (2) to the second law via Jensen's inequality. As was also said in the Introduction, a great deal of information is lost by this application of Jensen's inequality, while it is felt that eq. (2) has much more to tell.

In this section, we present a method by which a variety of stronger and more informative inequalities can be generated from eq. (2) by applying Jensen's inequality in a somewhat more sophisticated way, which allows a change of the probability measure, where the new measure is subjected to optimization. Unlike the usual use of Jensen's inequality, here no information is lost at all since eq. (2) can be 'recovered' as special case.

First, let us recall the elementary fact that the entropy production σ is actually a function of the random path (or the trajectory) $\boldsymbol{x} = \{x_t : \ 0 \leq t \leq T\}$ taken by the system in phase space as time runs from $t = 0$ to $t = T$, where x_t is the microscopic state at time t. In other words, σ should actually be denoted by $\sigma(\boldsymbol{x})$. For example, in the case of Markovian dynamics of a system controlled by an agent, $\sigma(\boldsymbol{x}) = \beta W_d(\boldsymbol{x}) = \beta[W(\boldsymbol{x}) - \Delta F]$, is the entropy production due to dissipated work $W_d(\boldsymbol{x})$ at fixed temperature $1/\beta$, where $W(\boldsymbol{x})$ is the total work and ΔF is the free–energy difference. This is then Jarzynski's equality.

Let the probability law, that governs the trajectory $\boldsymbol{x}$, be denoted by P. Let Q be an arbitrary alternative probability measure for $\boldsymbol{x}$. To avoid ambiguities, we denote, until further notice, expectations with respect to P and Q by $\langle\cdot\rangle_P$ and $\langle\cdot\rangle_Q$, respectively. Consider now the following chain of inequalities, where before applying Jensen's inequality, we change the underlying probability measure from P to Q, and thereby obtain an expression that depends on Q:

$$\begin{aligned} 1 &= \left\langle e^{-\sigma(\boldsymbol{x})} \right\rangle_P \\ &= \int \mathrm{d}\boldsymbol{x} P(\boldsymbol{x}) e^{-\sigma(\boldsymbol{x})} \\ &= \int \mathrm{d}\boldsymbol{x} Q(\boldsymbol{x}) e^{-\sigma(\boldsymbol{x}) + \ln[P(\boldsymbol{x})/Q(\boldsymbol{x})]} \end{aligned}$$

$$= \left\langle e^{-\sigma(\boldsymbol{x})+\ln[P(\boldsymbol{x})/Q(\boldsymbol{x})]} \right\rangle_Q$$
$$\geq e^{-\langle\sigma\rangle_Q - D(Q\|P)}. \qquad (64)$$

Here we have introduced the relative entropy $D(Q\|P) = \langle \ln[Q(\boldsymbol{x})/P(\boldsymbol{x})]\rangle_Q$. This is equivalent to the inequality

$$\langle\sigma\rangle_Q + D(Q\|P) \geq 0 \qquad (65)$$

which holds true for every probability measure Q. For $Q = P$, we are, of course, back to the second law. However, a stronger inequality is obtained upon minimizing the left hand side with respect to Q across some set of probability measures that includes $Q = P$. The global minimum of the left hand side among *all* probability measures (which is attained by $Q^*(\boldsymbol{x}) = P(\boldsymbol{x})e^{-\sigma(\boldsymbol{x})}$) turns out to be zero, and so, this leads to an uninteresting trivial identity. However, if the left hand side is minimized across some smaller set of measures (e.g., those that maintain moments of certain functions or physical quantities), then one still obtains a non–trivial inequality, which is stronger than the ordinary second law.

Another useful way to present inequality (65) is the following: Consider an arbitrary non–negative path function $\Psi = \Psi(\boldsymbol{x})$ with $\langle\Psi\rangle_P \in (0, \infty)$ and let us select

$$Q(\boldsymbol{x}) = \frac{P(\boldsymbol{x})\Psi(\boldsymbol{x})}{\int \mathrm{d}\boldsymbol{x}' P(\boldsymbol{x}')\Psi(\boldsymbol{x}')} = \frac{P(\boldsymbol{x})\Psi(\boldsymbol{x})}{\langle\Psi\rangle_P}. \qquad (66)$$

Then, obviously,

$$\langle\sigma\rangle_Q = \frac{\langle\Psi\cdot\sigma\rangle_P}{\langle\Psi\rangle_P}, \qquad (67)$$

and

$$D(Q\|P) = \int \mathrm{d}\boldsymbol{x}\cdot Q(\boldsymbol{x}) \log\frac{Q(\boldsymbol{x})}{P(\boldsymbol{x})}$$
$$= \int \mathrm{d}\boldsymbol{x}\cdot \frac{P(\boldsymbol{x})\Psi(\boldsymbol{x})}{\langle\Psi\rangle_P}\cdot\log\frac{\Psi(\boldsymbol{x})}{\langle\Psi\rangle_P}$$
$$= \frac{1}{\langle\Psi\rangle_P}\left(\langle\Psi\log\Psi\rangle_P - \langle\Psi\rangle_P \log\langle\Psi\rangle_P\right). \qquad (68)$$

On substituting these expressions back into (65), we obtain the following inequality, which is the main result of this section:

$$\langle\Psi\sigma\rangle \geq \langle\Psi\rangle\cdot\log\langle\Psi\rangle - \langle\Psi\log\Psi\rangle, \qquad (69)$$

where we have omitted the subscript P from the expectation operator since now, all expectations are taken again with respect to the original measure P. This extends the inequality $\langle\sigma\rangle \geq 0$ (the second law) by correlating $\sigma(\boldsymbol{x})$ with an arbitrary random variable $\Psi(\boldsymbol{x})$, measurable on the path. In other words, we now have a family of bounds on 'projections' of $\sigma(\boldsymbol{x})$ in the 'directions' of all non–negative path functions $\Psi(\boldsymbol{x})$, and not just in the one direction pertaining to $\Psi(\boldsymbol{x}) \equiv 1$, as in the second law. The right hand side of this inequality depends solely on the statistics of Ψ, and it is never positive due to the convexity of the function $f(t) = t\log t$ for $t \geq 0$.

The above lower bound is tight in the sense that there is a choice of $\Psi(\boldsymbol{x})$ for which the inequality becomes an equality, and this is $\Psi(\boldsymbol{x}) = e^{-\sigma(\boldsymbol{x})}$. Since the left hand side of the inequality is $\langle \sigma e^{-\sigma} \rangle$, for this choice of Ψ, and it is identical to the right hand side which is non–positive. It follows that

$$\left\langle \sigma e^{-\sigma} \right\rangle \leq 0. \tag{70}$$

Note that the last inequality can also be derived directly from the non-positivity of the right hand side of eq. (69). This means (similar to the usual interpretation of eq. (2)) that although σ is non–negative on the average, as the second law asserts, there is enough probabilistic weight to paths $\{\boldsymbol{x}\}$ for which $\sigma < 0$, so that the last inequality must hold.

A slightly more general case arises for $\Psi = e^{-\alpha\sigma}$ (α being an arbitrary real parameter), which yields

$$\left\langle \sigma e^{-\alpha\sigma} \right\rangle \geq \left\langle e^{-\alpha\sigma} \right\rangle \log \left\langle e^{-\alpha\sigma} \right\rangle + \alpha \left\langle \sigma e^{-\alpha\sigma} \right\rangle \ , \tag{71}$$

or equivalently,

$$\begin{aligned} \left\langle \sigma e^{-\alpha\sigma} \right\rangle &\geq \frac{\langle e^{-\alpha\sigma} \rangle \log \langle e^{-\alpha\sigma} \rangle}{1-\alpha} \qquad \alpha < 1 \\ \left\langle \sigma e^{-\alpha\sigma} \right\rangle &\leq \frac{\langle e^{-\alpha\sigma} \rangle \log \langle e^{-\alpha\sigma} \rangle}{1-\alpha} \qquad \alpha > 1 \ . \end{aligned} \tag{72}$$

Once again, the ordinary second law is obtained by substituting $\alpha = 0$ in the first of these two inequalities, which pertains to the case $\alpha < 1$.

Yet another interesting choice is $\Psi(\boldsymbol{x}) = \mathcal{I}\{\boldsymbol{x} : \ \sigma(\boldsymbol{x}) \leq \alpha\}$, where $\mathcal{I}\{\cdot\}$ is the indicator function of the event in the braces and α is an arbitrary real parameter. In this case, our inequality tells us something about the conditional expectation of σ given that $\sigma(\boldsymbol{x}) \leq \alpha$:

$$\langle \sigma \rangle \Big|_{\sigma \leq \alpha} \geq \log \Pr\{\sigma \leq \alpha\}, \tag{73}$$

where the ordinary second law is recovered as $\alpha \to \infty$. It may be interesting to look at the last inequality as an upper bounded on $\Pr\{\sigma \leq \alpha\}$:

$$\Pr\{\sigma \leq \alpha\} \leq \exp\left\{ \langle \sigma \rangle |_{\sigma \leq \alpha} \right\}, \tag{74}$$

which is interesting when $\langle \sigma \rangle |_{\sigma \leq \alpha} < 0$. For $\alpha < 0$, this is certainly the case. Note also that this is tighter than a straightforward use of the Chernoff bound [14, 15], according to

$$\Pr\{\sigma \leq \alpha\} \leq \left\langle e^{\alpha - \sigma} \right\rangle = e^{\alpha} \tag{75}$$

because $\langle \sigma \rangle |_{\sigma \leq \alpha}$ cannot exceed α.

5. Conclusion

In this paper, we have derived a series of bounds related to properties of the entropy production that stem from fluctuation theorems. These bounds illustrate that fluctuation theorems restrict the form of the probability distribution in a significant manner only for small mean entropy changes, but it is possible to improve these bounds upon incorporating additional information. Furthermore, we derived rigorous lower and upper bounds on the probability of the rare event where the entropy production is negative. The bounds depend on the mean entropy production and its variance, two quantities which are more readily accessible to measurement than the probability of this rare event. Finally, we presented a systematic way to derive inequalities which result from fluctuation theorems. These go beyond the standard derivation of the second law.

Acknowledgments: We are grateful for useful comments and discussions with Chris Jarzynski and Richard Blythe.

[1] For reviews see, for example, C. Bustamante, J. Liphardt and F. Ritort, Phys. Today **58**, 43 (2005); R. J. Harris, G. M. Schutz, J. Stat. Mech. – Theor. and Exp. P07020 (2007); J. Kurchan, J. Stat. Mech. – Theor. and Exp. P07005 (2007); Ritort F, S em. Poincar e **2**, 195 (2003); C. Maes, S em. ¶Poincar e **2** 29 (2003).

[2] For a review particularly relevant to this relation see D. J. Evans and D. J. Searles, Adv. Phys. **51**, 1529 (2002).

[3] For particularly simple derivations, see B. Cleuren, C. Van den Broeck and R. Kawai, Phys. Rev. Lett. **96**, 050601 (2006); U. Seifert, Phys. Rev. Letts. **95**, 040602 (2005); P. Pradhan, Y. Kafri and D. Levine, Phys. Rev. E, **77**, 041129 (2008).

[4] U. Seifert, Phys. Rev. Letts. **95**, 040602 (2005).

[5] D. J. Evans, E. G. D. Cohen and G. P. Morris, Phys. Rev. Lett. **71**, 2401 (1993). G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett. **74**, 2694 (1995).

[6] G. Gallavotti and E. G. D. Cohen, Phys. Rev. Lett., **74**, 2694 (1995).

[7] C. Jarzynski, and D. K. Wojcik, Phys. Rev. Lett. **92**, 230602 (2004).

[8] K. Saito, and A. Dhar, Phys. Rev. Lett. **99**, 180601 (2007).

[9] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[10] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[11] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).

[12] C. Jarzynski, J. Stat. Mech. – Theor. and Exp. P09005 (2004).

[13] G. E. Crooks, J. Stat. Phys. **90**, 14811 (1998).

[14] C. Jarzynski, Eur. Phys. J. B, **64**, 331 (2008).

[15] C. Jarzynski, J. Stat. Phys. **96**, 415 (1999).