CRAMER'S CONDITION AND SANOV'S THEOREM

Alexander Schied

Humboldt-Universität, D-10099 Berlin, Germany

Abstract

We discuss whether Sanov's theorem can be extended to a topology that renders the mapping $\nu \mapsto \int f\, d\nu$ continuous, for a given measurable function f. We show that this is possible if and only if f possesses all exponential moments with respect to the underlying law μ.

1. Introduction

Suppose $(E, \mathcal{E})$ is a measurable space, $M_1(E)$ is the space of all probability measures, and $\mu \in M_1(E)$ is fixed. If $\psi : E \to [1, \infty)$ is an $\mathcal{E}$-measurable function, we denote by $M_1^\psi(E)$ the set of all $\nu \in M_1(E)$ for which $\int \psi\, d\nu =: \langle \psi, \nu \rangle$ is finite. We endow $M_1^\psi(E)$ with the topology τ_ψ that is generated by the mappings

$$M_1^\psi(E) \ni \nu \mapsto \langle g, \nu \rangle, \qquad g \text{ is } \mathcal{E}\text{-measurable, and } |g| \le \psi.$$

The topology τ_1 is the so-called τ-topology on $M_1(E)$.

Now suppose $X_1, X_2, \dots$ is an sequence of independent E-valued random variables having the common law μ. Then Sanov's theorem states that the empirical distributions $\rho_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ satisfy a large deviation principle in $(M_1(E), \tau_1)$, with the good rate function $H(\cdot|\mu)$. Here $H(\nu|\mu)$ denotes the relative entropy of a measure ν with respect to μ:

$$H(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu}\, d\mu & \text{if } \nu \ll \mu, \\ \infty & \text{otherwise.} \end{cases}$$

See de Acosta (1994) for a proof of this result on an abstract measure space. Recently this theorem has been shown to hold also in the topological space $(M_1^\psi(E), \tau_\psi)$, provided the following *strong Cramér condition* holds:

(1) $$\int e^{\lambda\psi}\, d\mu < \infty, \qquad \text{for all } \lambda \in \mathbb{R}.$$

1991 Mathematics Subject Classification: 60F10

Key Words: Sanov's theorem, Cramér condition, Cramér's theorem, exponential moments

This is due to Wu (1993) in case where $(E, \mathcal{E})$ is a standard Borel space. See also Georgii and Zessin (1993) for related results. The general case was proved by Eichelsbacher and Schmock (1996). There even slightly stronger topologies are considered, which are induced by a whole collection Ψ of weight functions instead of a single ψ.

In this note, we will now discuss the question whether (1) can be replaced by the *weak Cramér condition*

$$\int e^{\delta\psi}\, d\mu < \infty, \qquad \text{for some } \delta > 0. \tag{2}$$

Indeed, one might conjecture that Sanov's theorem also holds under (2). So, for instance, the classical Cramér theorem states that (2) is a necessary and sufficient condition to have a large deviation principle with good rate function for the sequence $\psi(X_1), \psi(X_2), \dots$ (cf. Deuschel and Stroock (1989) or Dembo and Zeitouni (1993)). Also we will see below that (2) is necessary and sufficient in order that the level sets $\{\nu \mid H(\nu|\mu) \le \alpha\}$, $\alpha > 0$, are contained in $M_1^{\psi}(E)$. However, we will show that *compactness* of the level sets is equivalent to the *strong* Cramér condition (1). Therefore no reasonable analogue of Sanov's theorem holds in the space $(M_1^{\psi}(E), \tau_\psi)$ if only (2) is fulfilled.

Another motivation for studying this kind of question was the following. There has been recent interest in how to derive a large deviation principle on a locally convex topological vector space from a Sanov type result via contraction techniques. The reason for this is that in many cases the higher level result is much easier to derive. Such situations arise, for example, in the study of superprocesses and systems of interacting Markov processes. See, for instance, Section 1.8 of Fleischmann Gärtner and Kaj (1996) and the references therein, or Schied (1996). In many cases, this approach gives rise to the following question. Suppose we are given a μ-measurable function f on E. Then when is the mapping $\langle f, \cdot \rangle$ continuous on the level sets of the relative entropy $H(\cdot|\mu)$? The reason for this question is that continuity on the level sets is the minimal condition needed to apply a (generalized) contraction principle(cf. Dembo and Zeitouni (1993)). Below we will show that continuity of $\langle f, \cdot \rangle$ on the level sets of the entropy is in fact equivalent to a strong Cramér condition for f, which typically fails in the superprocess setting.

2. Results

Let $(E, \mathcal{E})$, μ, and ψ be as in Section 1. Then we have the following result.

Proposition 1: *Suppose $\alpha > 0$ is given. Then $\{\nu \mid H(\nu|\mu) \le \alpha\}$ is contained in $M_1^{\psi}(E)$ if and only if the weak Cramér condition (2) holds.*

However, our next theorem implies that Sanov's theorem cannot be extended to $M_1^{\psi}(E)$, if the strong Cramér condition (1) does not hold.

Theorem 2: *Suppose $\psi : E \to [1, \infty)$ is measurable, satisfies (2), and $\alpha > 0$ is given. Then the following are equivalent.*

1. *The level set $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$ is compact in $M_1^{\psi}(E)$.*
2. *The τ_1- and τ_ψ-topologies coincide on $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$.*
3. *The strong Cramér condition (1) holds: $\int e^{\lambda\psi}\, d\mu < \infty$, for all $\lambda \in \mathbb{R}$.*

Remark: If E is endowed with some Polish topology for which the Borel field is the same as $\mathcal{E}$, then the τ_1-continuity above can be replaced by weak continuity, because τ_1-topology and weak topology coincide on the level sets of the entropy.

Our Theorem 2 is in fact a corollary to the following result.

Theorem 3: *Suppose $f : E \to \mathbb{R}^+$ is a μ-measurable function. Then the following are equivalent.*

1. *The mapping $\langle f, \cdot\rangle$ is τ_1-continuous on $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$, for all $\alpha > 0$.*
2. *There is one $\alpha > 0$ such that $\langle f, \cdot\rangle$ is τ_1-continuous on $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$.*
3. *The strong Cramér condition holds for f: $\int e^{\lambda f}\, d\mu < \infty$, for all $\lambda \in \mathbb{R}$.*

The above theorem can easily be generalized to arbitrary μ-measurable functions $f : E \mapsto \mathbb{R}$, but one should assume that both f^+ and f^- satisfy the weak Cramér condition so that $\langle f, \cdot\rangle$ is well defined and finite on the level sets of the relative entropy.

The implications '$3 \Rightarrow 2$' of our Theorems 2 and 3 are already apparent in Wu (1993) and in Georgii and Zessin (1993). But below we will give a shorter proof of these assertions.

Our results show that even Cramér's theorem on $\mathbb{R}$ cannot be deduced in full generality by known contraction techniques from Sanov's theorem. This seems to be surprising, because one can show that the rate function I of Cramér's theorem has the form

$$I(m) = \inf \Big\{ H(\nu|\mu) \;\Big|\; \int x\, \nu(dx) = m \Big\}, \tag{3}$$

provided that we are in the regime where

$$\int e^{\lambda x}\, \mu(dx) < \infty \quad \Longleftrightarrow \quad \lambda^- < \lambda < \lambda^+,$$

for some $\lambda^- \in [-\infty, 0)$ and $\lambda^+ \in (0, \infty]$. Thus one would expect that there is a contraction principle behind (3). On the other hand, (3) may fail in an infinite dimensional setting, if only (2) and not (1) holds; see Lynch and Sethuraman (1987).

3. Proofs

Proof of Proposition 1: Recall the following classical variational identity for the relative entropy.

$$H(\nu|\mu) = \sup_{\phi \in B_b(E,\mathcal{E})} \Big(\int \phi \, d\nu - \log \int e^{\phi} \, d\mu \Big), \tag{4}$$

where $B_b(E, \mathcal{E})$ denotes the set of bounded $\mathcal{E}$-measurable functions on E. Monotone convergence implies that

$$H(\nu|\mu) + \log \int e^{\lambda\phi} \, d\mu \geq \lambda \cdot \int \phi \, d\nu \tag{5}$$

holds for all non-negative $\mathcal{E}$-measurable functions ϕ and all $\lambda \geq 0$. Now suppose (2) holds. Then we can choose $\phi = \psi$ and $\lambda = \delta$ in (5) to conclude that $H(\nu|\mu) < \infty$ implies that $\int \psi \, d\nu < \infty$.

Now suppose (2) fails. Then we can make the assumption that $\int \psi \, d\mu < \infty$, for otherwise the assertion would be trivial. Below we will construct a particular measure ν such that $0 < H(\nu|\mu) < \infty$ but $\int \psi \, d\nu = +\infty$. This will prove the assertion in case where $\alpha \geq H(\nu|\mu)$. If $\alpha > 0$ is arbitrary consider the line $\nu_t = (1-t)\mu + t\nu$, for $0 \leq t \leq 1$. Then $t \mapsto H(\nu_t|\mu)$ is a finite convex function, which hence is also continuous. Since in addition $H(\nu_0|\mu) = 0$, there is a $t_0 > 0$ such that $H(\nu_{t_0}|\mu) \leq \alpha$. To prove the existence of ν we will need the following simple lemma.

Lemma 4: *Suppose η is a (not necessarily finite) Borel measure on $[0, \infty)$ satisfying the following condition.*

(6) *For all $T > 0$ there is a Borel set $B_T \subset [T, \infty)$ such that $0 < \eta(B_T) < \infty$.*

Then there is a non-negative Borel measurable function g such that $\int g(x) \, \eta(dx) < \infty$ and $\int x g(x) \, \eta(dx) = +\infty$.

Proof: Suppose such a function would not exist. Then, if we define $Af(x) = xf(x)$, $x \in [0, \infty)$, we would have $Af \in L^1(\eta)$ for all $f \in L^1(\eta)$. Thus the linear operator A would have to be bounded by the Banach-Steinhaus theorem. However, if we denote the indicator of the set B_T by f_T we get $\int Af_T \, d\eta \geq T \int f_T \, d\eta > 0$. But this is a contradiction for T large enough. $\square$

Now we will construct the desired measure ν. By considering the distribution of ψ with respect to μ we may assume that $E = [0, \infty)$, $\mathcal{E}$ is the corresponding Borel field, and ψ is the identity map $\psi(x) = x$. If m denotes the function $m(x) = \mu([x, \infty))$ the failure of (2) implies that $\int_0^\infty e^{\lambda x} m(x)\, dx = \infty$, for all $\lambda > 0$. Hence $\ell(x) := -\log m(x)$ satisfies

$$\text{(7)} \qquad \int_0^\infty \mathrm{I}_{\{(\ell(x)+1)/x < \varepsilon\}}\, dx > 0 \qquad \text{for all } \varepsilon > 0.$$

Choose a function a taking values in $[1, 2]$ such that the distribution η of $(a(x)+x)/(1+\ell(x))$ with respect to the measure $(1+x)^{-1}\, dx$ does not have atoms of infinite mass. Because of (7), η satisfies assumption (6). Lemma 4 hence gives us the existence of a non-negative function g such that

$$\text{(8)} \quad \int_0^\infty g\Big(\frac{a(x)+x}{\ell(x)+1}\Big)\frac{1}{1+x}\, dx < \infty \ \text{ and } \int_0^\infty g\Big(\frac{a(x)+x}{\ell(x)+1}\Big)\frac{a(x)+x}{(\ell(x)+1)(1+x)}\, dx = +\infty.$$

We will write for short $\rho(x)$ for the integrand in the second integral of (8). Clearly

$$c^{-1} := \int \int_0^x \frac{\rho(y)}{m(y)(1+y)}\, dy\, \mu(dx) = \int_0^\infty \frac{\rho(y)}{1+y}\, dy < \infty,$$

so that

$$\nu(dx) := c \cdot \int_0^x \frac{\rho(y)}{m(y)(1+y)}\, dy\, \mu(dx)$$

is a probability measure. Of course $\nu \neq \mu$ and hence $0 < H(\nu|\mu)$. By (8) it follows that

$$\int x\, \nu(dx) = \int_0^\infty m(x)\frac{d\nu}{d\mu}(x) + \frac{c \cdot \rho(x)}{1+x} x\, dx = +\infty$$

On the other hand, a similar calculation shows that

$$H(\nu|\mu) = 1 + c \int_0^\infty \frac{\rho(x)}{1+x} \log \frac{d\nu}{d\mu}(x)\, dx.$$

But $d\nu/d\mu(x) \leq m(x)^{-1}$ and hence

$$H(\nu|\mu) \leq 1 + c \int_0^\infty \frac{\rho(x)}{1+x} \ell(x)\, dx < \infty$$

again by (8). Thus we have constructed the desired measure ν, and the assertion of Proposition 1 is proved. $\square$

The proof of the remaining two theorems relies on the following simple lemma. Its implication '$\Leftarrow$' is already implicitly contained in Georgii and Zessin (1993).

Lemma 5: *Suppose $K \subset M_1^+(E)$ is τ_1-compact, and $f \geq 0$is a measurable function. Then $\langle f, \cdot \rangle$ is τ_1-continuous on K, if and only if the following condition is fulfilled.*

(9) *For any $\varepsilon > 0$ one can find $c > 0$ such that* $\sup_{\nu \in K} \int_{\{f \geq c\}} f \, d\nu \leq \varepsilon.$

Proof: Suppose $\langle f, \cdot \rangle$ is τ_1-continuous on K. Then so is

$$\nu \mapsto \int_{\{f \geq c\}} f \, d\nu = \int f \, d\nu - \int_{\{f < c\}} f \, d\nu.$$

But $\int_{\{f \geq c\}} f \, d\nu$ decreases to 0 as $c \uparrow \infty$, for any $\nu \in K$, and this convergence is even uniform on K due to Dini's theorem. This implies (9).

On the other hand, (9) implies that

$$\sup_{\nu \in K} \left| \int f \, d\nu - \int_{\{f < c\}} f \, d\nu \right| = \sup_{\nu \in K} \int_{\{f \geq c\}} f \, d\nu \longrightarrow 0 \qquad \text{as } c \uparrow \infty.$$

Thus $\langle f, \cdot \rangle$ is a uniform limit of continuous functions on K. $\square$

Recall the fact that $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$ is a τ_1-compact set for each $\alpha \geq 0$ (c.f. Deuschel and Stroock (1989) or Dembo and Zeitouni (1993)).

Proof of Theorem 3: First we will show the implication $2 \Rightarrow 3$ of Theorem 3. To this end, let us assume that $\langle f, \cdot \rangle$ is τ_1-continuous on $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$ for some $\alpha > 0$, but that

$$\lambda^* := \sup \left\{ \lambda \in \mathbb{R} \,\Big|\, \int e^{\lambda f} \, d\mu < \infty \right\} < \infty.$$

If $\lambda^* = 0$ we already know from Proposition 1 that the functional $\langle f, \cdot \rangle$ is unbounded on $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$. Thus it cannot be continuous. Therefore we can assume $\lambda^* > 0$ in the sequel. For $\lambda \geq 0$, $c > 0$, and $s > c$ let

$$\Lambda_s(\lambda) = \log \int \exp\left(\lambda f \wedge s \cdot \mathrm{I}_{\{f \geq c\}} \right) d\mu,$$

and define probability measures ν_λ^s by

$$d\nu_\lambda^s = \exp\left(\lambda f \wedge s \cdot \mathrm{I}_{\{f \geq c\}} - \Lambda_s(\lambda) \right) d\mu,$$

Then Λ_s is non-negative, smooth, and convex as a function of λ. Moreover, it is well-known that

$$H(\nu_\lambda^s|\mu) = \lambda \frac{d}{d\lambda}\Lambda_s(\lambda) - \Lambda_s(\lambda). \tag{10}$$

In particular we get the inequality

$$\int_{\{f\geq c\}} f \wedge s \, d\nu_\lambda^s = \frac{d}{d\lambda}\Lambda_s(\lambda) \geq \frac{1}{\lambda} H(\nu_\lambda^s|\mu). \tag{11}$$

Now fix $\lambda' \in (\lambda^*, 2\lambda^*)$. Note that $\Lambda_s(\lambda) \to \Lambda_\infty(\lambda)$ as $s \uparrow \infty$, with $\Lambda_\infty(\lambda') = \infty$ for all c, while $\Lambda_\infty(\lambda'/2)$ is finite. Equation (10) and the convexity of Λ_s imply that also

$$H(\nu_{\lambda'}^s|\mu) \geq \Lambda_s(\lambda') - 2\Lambda_s(\lambda'/2) \longrightarrow \infty \qquad \text{as } s \uparrow \infty.$$

On the other hand, it follows from (10) that $H(\nu_\lambda^s|\mu)$ is a continuous (and even smooth) function of λ. Hence, for s large enough, there exists $\lambda_s \in (0, \lambda')$ such that $H(\nu_{\lambda_s}^s|\mu) = \alpha$. Thus we get from (11) that

$$\int_{\{f\geq c\}} f \, d\nu_{\lambda_s}^s \geq \int_{\{f\geq c\}} f \wedge s \, d\nu_{\lambda_s}^s \geq \frac{\alpha}{\lambda_s} \geq \frac{\alpha}{\lambda'}$$

holds for all c and s large enough, and for $\lambda' \in (\lambda^*, 2\lambda^*)$. Hence

$$\sup_{\nu : H(\nu|\mu)\leq\alpha} \int_{\{f\geq c\}} f \, d\nu \geq \frac{\alpha}{\lambda^*}, \qquad \text{for all } c > 0,$$

and condition (9) is violated. Therefore $\langle f, \cdot\rangle$ cannot be continuous on $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$.

In order to prove the implication $3 \Rightarrow 1$ let $\lambda = 2\alpha/\varepsilon$ and choose $c > 0$ such that

$$\log \int \exp\left(\lambda f \cdot \mathrm{I}_{\{f\,\geq\,c\}}\right) d\mu \leq \alpha.$$

Then taking $\phi = f \cdot \mathrm{I}_{\{f\geq c\}}$ in (5) yields $\int_{\{f\geq c\}} f \, d\nu \leq \varepsilon$, for all ν with $H(\nu|\mu) \leq \alpha$. Hence $\langle f, \cdot\rangle$ is τ_1-continuous on $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$ by Lemma 5, and Theorem 3 is proved. □

Proof of Theorem 2: Let K_α denote the τ_1-compact level set $\{\nu \,|\, H(\nu|\mu) \leq \alpha\}$. First we show $1 \iff 2$. Indeed, since the identity is a continuous and bijective mapping from (K_α, τ_ψ) to (K_α, τ_1), it is also a homeomorphism if and only if K_α is τ_ψ-compact. Now we turn to the proof of $2 \iff 3$: It follows from Theorem 1 that every function $\langle g, \cdot\rangle$ with $|g| \leq \psi$ is τ_1-continuous on K_α, if and only if 3. holds. □

Acknowledgment. I would like to thank the Associate Editor and the anonymous referee of this paper for valuable comments and for indicating a simplification of the original argument in the proof of Theorem 3. Furthermore I am grateful to J. Brettschneider, P. Eichelsbacher, T. Strobel, and, in particular, U. Schmock for many remarks which greatly improved previous versions of the manuscript.

References

de Acosta, A. (1994), On large deviations of empirical measures in the τ-topology. Studies in applied probability. J. Appl. Probab. **31A** (special volume), 41-47.

Dembo, A., Zeitouni, O. (1993), *Large deviation techniques* (Jones and Bartlett Publishers, Boston, London).

Deuschel, J. D., Stroock, D. W. (1989), *Large deviations* (Academic Publishers, Boston).

Eichelsbacher, P., Schmock, U. (1996), Large deviations for empirical processes and U-empirical measures in strong topologies, Preprint Universität Bielefeld.

Fleischmann, K., Gärtner J., Kaj I. (1996), A Schilder type theorem for super-Brownian motion, Canadian J. Math. **48**, 542-568.

Georgii, H. O., Zessin, H. (1993) Large deviations and the maximum entropy principle for marked point random fields, Probab. Theory Relat. Fields, **96**, 177-204.

Lynch, J., Sethuraman, J. (1987), Large deviations for processes with independent increments Ann. Probab. **15**, 610-627.

Schied, A. (1996), Sample path large deviations for super-Brownian motion, Probab. Theory Related Fields **104**, 319-347.

Wu, L. M. (1993), Some general methods of large deviations and applications, Preprint.