

# Subgaussian random variables, Hoeffding's inequality, and Cramér's large deviation theorem

Jordan Bell

June 4, 2014

## 1 Subgaussian random variables

For a random variable  $X$ , let  $\Lambda_X(t) = \log E(e^{tX})$ , the **cumulant generating function of  $X$** . A  **$b$ -subgaussian random variable**,  $b > 0$ , is a random variable  $X$  such that

$$\Lambda_X(t) \leq \frac{b^2 t^2}{2}, \quad t \in \mathbb{R}.$$

We remark that for  $\gamma_{a,\sigma^2}$  a **Gaussian measure**, whose density with respect to Lebesgue measure on  $\mathbb{R}$  is

$$p(x, a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

we have

$$\int_{\mathbb{R}} e^{tx} d\gamma_{0,b^2}(x) = \int_{\mathbb{R}} e^{bty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \int_{\mathbb{R}} e^{\frac{b^2 t^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-bt)^2}{2}} dy = e^{\frac{b^2 t^2}{2}}.$$

We prove that a  $b$ -subgaussian random variable is centered and has variance  $\leq b^2$ .<sup>1</sup>

**Theorem 1.** *If  $X$  is  $b$ -subgaussian then  $E(X) = 0$  and  $\text{Var}(X) \leq b^2$ .*

*Proof.* For each  $\omega \in \Omega$ ,  $\sum_{k=0}^n \frac{t^k X(\omega)^k}{k!} \rightarrow e^{tX(\omega)}$ , and by the dominated convergence theorem,

$$\sum_{k=0}^n \frac{t^k E(X)^k}{k!} \rightarrow E(e^{tX}) \leq e^{\frac{b^2 t^2}{2}} = \sum_{k=0}^{\infty} \left( \frac{b^2 t^2}{2} \right)^k \frac{1}{k!}.$$

Therefore

$$1 + tE(X) + t^2 E(X^2) + O(t^3) \leq 1 + \frac{b^2 t^2}{2} + O(t^4),$$

---

<sup>1</sup>Karl R. Stromberg, *Probability for Analysts*, p. 293, Proposition 9.8.

whence

$$tE(X) + t^2E(X^2) \leq \frac{b^2t^2}{2} + o(t^2),$$

and so, for  $t > 0$ ,

$$E(X) + tE(X^2) \leq \frac{b^2t}{2} + o(t).$$

First, this yields  $E(X) = o(t)$ , which means that  $E(X) = 0$ . Second, since  $E(X) = 0$ ,

$$tE(X^2) \leq \frac{b^2t}{2} + o(t),$$

and then

$$E(X^2) \leq \frac{b^2}{2} + o(1),$$

which means that  $E(X^2) \leq \frac{b^2}{2}$ . □

Stromberg attributes the following theorem to Saeki; further, it is proved in Stromberg that if for some  $t$  the inequality in the theorem is an equality then the random variable has the Rademacher distribution.<sup>2</sup>

**Theorem 2.** *If  $X$  is a random variable satisfying  $E(X) = 0$  and  $P(X \in [-1, 1]) = 1$ , then*

$$E(e^{tX}) \leq \cosh t, \quad t \in \mathbb{R}.$$

*Proof.* Define  $f : \mathbb{R} \rightarrow \mathbb{R}$  by

$$f(t) = e^t (\cosh t - E(e^{tX})) = \frac{e^{2t}}{2} + \frac{1}{2} - e^t E(e^{tX}).$$

Then

$$f'(t) = e^{2t} - e^t E(e^{tX}) - e^t E(Xe^{tX});$$

the derivative of  $E(e^{tX})$  with respect to  $t$  is obtained using the dominated convergence theorem. Let  $Y = 1 + X$ , with which

$$f'(t) = e^{2t} - E(e^{tY}) - E(Xe^{tY}) = e^{2t} - E(e^{tY}) - E((Y-1)e^{tY}) = e^{2t} - E(Ye^{tY}).$$

$E(X) = 0$ , so  $E(Y) = 1$ , hence

$$f'(t) = E(e^{2t}Y) - E(Ye^{tY}) = E(Y(e^{2t} - e^{tY})).$$

Because  $P(Y \in [0, 2]) = 1$ , for  $t \geq 0$ , we have almost surely  $e^{2t} - e^{tY} \geq 0$ , and therefore almost surely  $Y(e^{2t} - e^{tY}) \geq 0$ . Therefore, for  $t \geq 0$ ,

$$f'(t) = E(Y(e^{2t} - e^{tY})) \geq 0,$$

---

<sup>2</sup>Karl R. Stromberg, *Probability for Analysts*, p. 293, Proposition 9.9; Omar Rivasplata, *Subgaussian random variables: An expository note*, <http://www.math.ualberta.ca/~orivasplata/publications/subgaussians.pdf>

which tells us that for  $t \geq 0$ ,

$$f(0) \leq f(t).$$

As  $f(0) = 0$ , for  $t \geq 0$ ,

$$0 \leq e^t (\cosh t - E(e^{tX})) ,$$

and so

$$E(e^{tX}) \leq \cosh t.$$

□

**Corollary 3.** *If a random variable  $X$  satisfies  $E(X) = 0$  and  $P(|X| \leq b) = 1$ , then  $X$  is  $b$ -subgaussian.*

## 2 Hoeffding's inequality

We first prove **Hoeffding's lemma**.<sup>3</sup>

**Lemma 4** (Hoeffding's lemma). *If a random variable  $X$  satisfies  $E(X) = 0$  and  $P(X \in [a, b]) = 1$ , then  $X$  is  $\frac{b-a}{2}$ -subgaussian.*

*Proof.* Because  $P(X \in [a, b]) = 1$ , it follows that

$$\text{Var}(X) \leq \frac{(b-a)^2}{4},$$

not using that  $P(X) = 0$ . (Namely, Popoviciu's inequality.)

Write  $\mu = X_*P$  and for  $\lambda \in \mathbb{R}$  define

$$d\nu_\lambda(t) = \frac{e^{\lambda t}}{e^{\Lambda(\lambda)}} d\mu(t).$$

We check

$$\int_{\mathbb{R}} d\nu_\lambda(t) = \frac{1}{e^{\Lambda(\lambda)}} \int_{\mathbb{R}} e^{\lambda t} d(X_*P)(t) = \frac{1}{e^{\Lambda(\lambda)}} \int_{\Omega} e^{\lambda X} dP = 1.$$

There is a random variable  $X_\lambda : (\Omega_\lambda, \mathcal{F}_\lambda, P_\lambda) \rightarrow \mathbb{R}$  for which  $X_{\lambda*}P_\lambda = \nu_\lambda$ .  $X_\lambda$  satisfies  $P_\lambda(X_\lambda \in [a, b]) = 1$ , and so

$$\text{Var}(X_\lambda) \leq \frac{(b-a)^2}{4}.$$

We calculate

$$\Lambda'_X(t) = \frac{E(X e^{tX})}{E(e^{tX})}$$

---

<sup>3</sup>Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, p. 27, Lemma 2.2.

and

$$\Lambda_X''(t) = \frac{E(X^2 e^{tX})E(e^{tX}) - E(Xe^{tX})E(Xe^{tX})}{E(e^{tX})^2}.$$

But

$$E(X_\lambda) = \int_{\mathbb{R}} t d\nu_\lambda(t) = \int_{\mathbb{R}} t \frac{e^{\lambda t}}{e^{\Lambda(\lambda)}} d\mu(t) = \frac{1}{e^{\Lambda(\lambda)}} E(Xe^{\lambda X})$$

and

$$E(X_\lambda^2) = \int_{\mathbb{R}} t^2 d\nu_\lambda(t) = \frac{1}{e^{\Lambda(\lambda)}} E(X^2 e^{\lambda X}),$$

and so

$$\begin{aligned} \text{Var}(X_\lambda) &= E(X_\lambda^2) - E(X_\lambda)^2 \\ &= \frac{E(X^2 e^{\lambda X})}{e^{\Lambda(\lambda)}} - \frac{E(Xe^{\lambda X})^2}{e^{2\Lambda(\lambda)}} \\ &= \Lambda_X''(\lambda). \end{aligned}$$

For  $\lambda \in \mathbb{R}$ , Taylor's theorem tells us that there is some  $\theta$  between 0 and  $\lambda$  such that

$$\Lambda_X(\lambda) = \Lambda_X(0) + \lambda \Lambda_X'(0) + \frac{\lambda^2}{2} \Lambda_X''(\theta) = \frac{\lambda^2}{2} \Lambda_X''(\theta);$$

here we have used that  $E(X) = 0$ . But from what we have shown,  $\text{Var}(X_\theta) = \Lambda_X''(\theta)$  and  $\text{Var}(X_\theta) \leq \frac{(b-a)^2}{4}$ , so

$$\Lambda_X(\lambda) = \frac{\lambda^2}{2} \text{Var}(X_\theta) \leq \frac{\lambda^2}{2} \cdot \frac{(b-a)^2}{4},$$

which shows that  $X$  is  $\frac{b-a}{2}$ -subgaussian.  $\square$

We now prove **Hoeffding's inequality**.<sup>4</sup>

**Theorem 5** (Hoeffding's inequality). *Let  $X_1, \dots, X_n$  be independent random variables such that for each  $1 \leq k \leq n$ ,  $P(X_k \in [a_k, b_k]) = 1$ , and write  $S_n = \sum_{k=1}^n X_k$ . For any  $a > 0$ ,*

$$P(S_n - E(S_n) \geq a) \leq \exp\left(-\frac{2a^2}{\sum_{k=1}^n (b_k - a_k)^2}\right).$$

*Proof.* For  $\lambda > 0$  and  $\phi(t) = e^{\lambda t}$ , because  $\phi$  is nonnegative and nondecreasing, for  $X$  a random variable we have

$$1_{X \geq a} \phi(a) \leq \phi(X),$$

and so  $E(1_{X \geq a} \phi(a)) \leq E(\phi(X))$ , i.e.

$$P(X \geq a) \leq \frac{E(e^{\lambda X})}{e^{\lambda a}}.$$

---

<sup>4</sup>Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*, p. 34, Theorem 2.8.

Using this with  $X = S_n - E(S_n)$  and because the  $X_k$  are independent,

$$P(S_n - E(S_n) \geq a) \leq \frac{1}{e^{\lambda a}} E(e^{\lambda(S_n - E(S_n))}) = e^{-\lambda a} \prod_{k=1}^n E(e^{\lambda(X_k - E(X_k))}).$$

Because  $P(X_k \in [a_k, b_k]) = 1$ , we have  $P(X_k - E(X_k) \in [a_k - E(X_k), b_k - E(X_k)]) = 1$ , and as  $(b_k - E(X_k)) - (a_k - E(X_k)) = b_k - a_k$ , Hoeffding's lemma tells us

$$\log E(e^{\lambda(X_k - E(X_k))}) \leq \frac{(b_k - a_k)^2 \lambda^2}{8},$$

and thus

$$\begin{aligned} P(S_n - E(S_n) \geq a) &\leq e^{-\lambda a} \exp\left(\sum_{k=1}^n \frac{(b_k - a_k)^2 \lambda^2}{8}\right) \\ &= \exp\left(-\lambda a + \frac{\lambda^2}{8} \sum_{k=1}^n (b_k - a_k)^2\right). \end{aligned}$$

We remark that  $\lambda$  does not appear in the left-hand side. Define

$$g(\lambda) = -\lambda a + \frac{\lambda^2}{8} \sum_{k=1}^n (b_k - a_k)^2,$$

for which

$$g'(\lambda) = -a + \frac{\lambda}{4} \sum_{k=1}^n (b_k - a_k)^2.$$

Then  $g'(\lambda) = 0$  if and only if

$$\lambda = \frac{4a}{\sum_{k=1}^n (b_k - a_k)^2},$$

at which  $g$  assumes its infimum. Then

$$\begin{aligned} P(S_n - E(S_n) \geq a) &\leq \exp\left(-\frac{4a^2}{\sum_{k=1}^n (b_k - a_k)^2} + \frac{16a^2}{8} \frac{1}{\sum_{k=1}^n (b_k - a_k)^2}\right) \\ &= \exp\left(-\frac{2a}{\sum_{k=1}^n (b_k - a_k)^2}\right), \end{aligned}$$

proving the claim. □

### 3 Cramér's large deviation theorem

The following is **Cramér's large deviation theorem**.<sup>5</sup>

---

<sup>5</sup>Achim Klenke, *Probability Theory: A Comprehensive Course*, p. 508, Theorem 23.3.

**Theorem 6** (Cramér's large deviation theorem). *Suppose that  $X_n : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{R}$ ,  $n \geq 1$ , are independent identically distributed random variables such that for all  $t \in \mathbb{R}$ ,*

$$\Lambda(t) = \log E(e^{tX_1}) < \infty.$$

For  $x \in \mathbb{R}$  define

$$\Lambda^*(x) = \sup_{t \in \mathbb{R}} (tx - \Lambda(t)).$$

If  $a > E(X_1)$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq an) = -\Lambda^*(a),$$

where  $S_n = \sum_{k=1}^n X_k$ .

*Proof.* For  $a > E(X_1)$ , let  $Y_n = X_n - a$ , let

$$L(t) = \log E(e^{tY_1}) = \log E(e^{tX_1} e^{-ta}) = -ta + \Lambda(t)$$

and let

$$L^*(x) = \sup_{t \in \mathbb{R}} (tx - L(t)) = \sup_{t \in \mathbb{R}} (t(x + a) - \Lambda(t)) = \Lambda^*(x + a).$$

Lastly, let  $T_n = \sum_{k=1}^n Y_k = S_n - na$ , with which

$$P(T_n \geq bn) = P(S_n \geq (b + a)n).$$

Thus, if we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(T_n \geq 0) = -L^*(0),$$

then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq an) = -L^*(0) = -\Lambda^*(a).$$

Therefore it suffices to prove the theorem for when  $E(X_1) < 0$  and  $a = 0$ .

Define

$$\phi(t) = e^{\Lambda(t)} = E(e^{tX_1}) = \int_{\Omega} e^{tX_1} dP = \int_{\mathbb{R}} e^{tx} d(X_{1*}P)(x), \quad t \in \mathbb{R},$$

the moment generating function of  $X_1$ , and define

$$\rho = e^{-\Lambda^*(0)} = \exp \left( -\sup_{t \in \mathbb{R}} (-\Lambda(t)) \right) = \exp \left( \inf_{t \in \mathbb{R}} \Lambda(t) \right) = \inf_{t \in \mathbb{R}} \phi(t),$$

using that  $x \mapsto e^x$  is increasing.

Using the dominated convergence theorem, for  $k \geq 0$  we obtain

$$\phi^{(k)}(t) = \int_{\mathbb{R}} x^k e^{tx} d(X_{1*}P)(x) = E(X_1^k e^{tX_1}).$$

In particular,  $\phi'(t) = E(X_1 e^{tX_1})$ , for which  $\phi'(0) = E(X_1) < 0$ , and  $\phi''(t) = E(X_1^2 e^{tX_1}) > 0$  for all  $t$  (either the expectation is 0 or positive, and if it is 0 then  $X_1^2 e^{tX_1}$  is 0 almost everywhere, which contradicts  $E(X_1) < 0$ ).

Either  $P(X_1 \leq 0) = 1$  or  $P(X_1 \leq 0) < 1$ . In the first case,

$$\phi'(t) = \int_{\Omega} X_1 e^{tX_1} dP = \int_{X_1 \leq 0} X_1 e^{tX_1} dP \leq 0,$$

so, using the dominated convergence theorem,

$$\rho = \inf_{t \in \mathbb{R}} \phi(t) = \lim_{t \rightarrow \infty} \phi(t) = \int_{X_1 \leq 0} \left( \lim_{t \rightarrow \infty} e^{tX_1} \right) dP = \int_{X_1=0} dP = P(X_1 = 0).$$

Then

$$P(S_n \geq 0) = P(X_1 = 0, \dots, X_n = 0) = P(X_1 = 0) \cdots P(X_n = 0) = \rho^n.$$

That is, as  $a = 0$ ,

$$P(S_n \geq a) = e^{-n\Lambda^*(a)},$$

and the claim is immediate in this case.

In the second case,  $P(X_1 \leq 0) < 1$ . As  $\phi''(t) > 0$  for all  $t$ , there is some  $\tau \in \mathbb{R}$  at which  $\phi(\tau) < \phi(t)$  for all  $t \neq \tau$  (namely, a unique global minimum). Thus,

$$\phi(\tau) = \rho, \quad \phi'(\tau) = 0.$$

And  $\phi'(0) = E(X_1) < 0$ , which with the above yields  $\tau > 0$ . Because  $\tau > 0$ ,  $S_n(\omega) \geq 0$  if and only if  $\tau S_n(\omega) \geq 0$  if and only if  $e^{\tau S_n(\omega)} \geq 1$ . Applying Chebyshev's inequality, and because  $X_n$  are independent,

$$P(S_n \geq 0) = P(e^{\tau S_n} \geq 1) \leq E(e^{\tau S_n}) = E(e^{\tau X_1}) \cdots E(e^{\tau X_n}) = \phi(\tau)^n = \rho^n,$$

thus  $\log P(S_n \geq 0) \leq n \log \rho$  and then

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq 0) \leq \limsup_{n \rightarrow \infty} \log \rho = \log \rho = \log e^{-\Lambda^*(0)} = -\Lambda^*(0).$$

To prove the claim, it now suffices to prove that, in the case  $P(X_1 \leq 0) < 1$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq 0) \geq \log \rho. \quad (1)$$

Let  $\mu = X_{1*}P$ , and let

$$d\nu(x) = \frac{e^{\tau x}}{\rho} d\mu(x).$$

$\nu$  is a Borel probability measure: it is apparent that it is a Borel measure, and

$$\nu(\mathbb{R}) = \int_{\mathbb{R}} d\nu(x) = \int_{\mathbb{R}} \frac{e^{\tau x}}{\rho} d\mu(x) = \frac{1}{\rho} \int_{\mathbb{R}} e^{\tau x} d\mu(x) = \frac{\phi(\tau)}{\rho} = 1.$$

There are independent identically distributed random variables  $Y_n$ ,  $n \geq 1$ , each with  $Y_n * Q = \nu$ .<sup>6</sup> Define

$$\psi(t) = E(e^{tY_1}) = \int_{\mathbb{R}} e^{tx} d\nu(x) = \int_{\mathbb{R}} e^{tx} \frac{e^{\tau x}}{\rho} d\mu(x) = \frac{1}{\rho} \int_{\mathbb{R}} e^{(t+\tau)x} d\mu(x) = \frac{\phi(t+\tau)}{\rho},$$

the moment generating function of  $Y_1$ . As  $\phi'(\tau) = 0$ ,

$$E(Y_1) = \psi'(0) = \frac{\phi'(\tau)}{\rho} = 0.$$

As  $\rho > 0$  and  $\phi''(t) > 0$  for all  $t$ ,

$$\text{Var}(Y_1) = E(Y_1^2) = \psi''(0) = \frac{\phi''(\tau)}{\rho} \in (0, \infty).$$

For  $T_n = \sum_{k=1}^n Y_k$ , using that the  $X_n$  are independent and that the  $Y_n$  are independent,

$$\begin{aligned} P(S_n \geq 0) &= \int_{x_1 + \dots + x_n \geq 0} d(S_n * P)(x) \\ &= \int_{x_1 + \dots + x_n \geq 0} d\mu(x_1) \cdots d\mu(x_n) \\ &= \int_{x_1 + \dots + x_n \geq 0} \left( \frac{\rho}{e^{\tau x_1}} d\nu(x_1) \right) \cdots \left( \frac{\rho}{e^{\tau x_n}} d\nu(x_n) \right) \\ &= \rho^n \int_{x_1 + \dots + x_n \geq 0} e^{-\tau(x_1 + \dots + x_n)} d(T_n * Q). \end{aligned}$$

But

$$\begin{aligned} \int_{x_1 + \dots + x_n \geq 0} e^{-\tau(x_1 + \dots + x_n)} d(T_n * Q) &= \int_{T_n \geq 0} e^{-\tau T_n} dQ \\ &= E(1_{\{T_n \geq 0\}} \cdot e^{-\tau T_n}), \end{aligned}$$

hence

$$P(S_n \geq 0) = \rho^n E(1_{\{T_n \geq 0\}} \cdot e^{-\tau T_n}).$$

Thus, (1) is equivalent to

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log (\rho^n E(1_{\{T_n \geq 0\}} \cdot e^{-\tau T_n})) \geq \log \rho,$$

so, to prove the claim it suffices to prove that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log (E(1_{\{T_n \geq 0\}} \cdot e^{-\tau T_n})) \geq 0.$$

---

<sup>6</sup>Gerald B. Folland, *Real Analysis: Modern Techniques and Their Applications*, p. 329, Corollary 10.19.



For any  $c > 0$ ,

$$\begin{aligned}
\log (E(1_{\{T_n \geq 0\}} \cdot e^{-\tau T_n})) &\geq \log E(1_{\{0 \leq T_n \leq c\sqrt{n}\}} \cdot e^{-\tau T_n}) \\
&\geq \log (e^{-\tau c\sqrt{n}} \cdot Q(0 \leq T_n \leq c\sqrt{n})) \\
&= -\tau c\sqrt{n} + \log Q\left(\frac{T_n}{\sqrt{n}} \in [0, c]\right).
\end{aligned}$$

Because the  $Y_n$  are independent identically distributed  $L^2$  random variables with mean 0 and variance  $\sigma^2 = \text{Var}(Y_1) = \frac{\phi''(\tau)}{\rho}$ , the central limit theorem tells us that as  $n \rightarrow \infty$ ,

$$Q\left(\frac{T_n}{\sqrt{n}} \in [0, c]\right) \rightarrow \gamma_{0, \sigma^2}([0, c]),$$

where  $\gamma_{a, \sigma^2}$  is the Gaussian measure, whose density with respect to Lebesgue measure on  $\mathbb{R}$  is

$$p(t, a, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}}.$$

Thus, because for  $c > 0$  we have  $\gamma_{0, \sigma^2}([0, c]) > 0$ ,

$$\begin{aligned}
\liminf_{n \rightarrow \infty} \frac{1}{n} \log (E(1_{\{T_n \geq 0\}} \cdot e^{-\tau T_n})) &\geq \liminf_{n \rightarrow \infty} \left( \frac{-\tau c}{\sqrt{n}} + \frac{1}{n} \log Q\left(\frac{T_n}{\sqrt{n}} \in [0, c]\right) \right) \\
&= \lim_{n \rightarrow \infty} -\frac{\tau c}{\sqrt{n}} + \lim_{n \rightarrow \infty} \frac{1}{n} \log Q\left(\frac{T_n}{\sqrt{n}} \in [0, c]\right) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \log \gamma_{0, \sigma^2}([0, c]) \\
&= 0,
\end{aligned}$$

which completes the proof.  $\square$

For example, say that  $X_n$  are independent identically distributed random variables with  $X_{1*}P = \gamma_{0,1}$ . We calculate that the cumulant generating function  $\Lambda(t) = \log E(e^{tX_1})$  is

$$\begin{aligned}
\Lambda(t) &= \log \left( \int_{\mathbb{R}} e^{tx} d\gamma_{0,1}(x) \right) \\
&= \log \left( \int_{\mathbb{R}} e^{tx} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \right) \\
&= \log \left( \int_{\mathbb{R}} \frac{e^{-\frac{1}{2}(x-t)^2}}{\sqrt{2\pi}} e^{\frac{t^2}{2}} dx \right) \\
&= \log e^{\frac{t^2}{2}} 2 \\
&= \frac{t^2}{2},
\end{aligned}$$

thus  $\Lambda(t) < \infty$  for all  $t$ . Then

$$\Lambda^*(x) = \sup_{t \in \mathbb{R}} (tx - \Lambda(t)) = \sup_{t \in \mathbb{R}} \left( tx - \frac{t^2}{2} \right) = \frac{x^2}{2}.$$

Now applying Cramér's theorem we get that for  $a > E(X_1) = 0$ , for  $S_n = \sum_{k=1}^n X_k$  we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq an) = -\frac{a^2}{2}.$$

Another example: If  $X_n$  are independent identically distributed random variables with the **Rademacher distribution**:

$$X_n * P = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_1.$$

Then

$$E(e^{tX_1}) = \int_{\mathbb{R}} e^{tx} d\left(\frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_1\right)(x) = \frac{1}{2} e^{-t} + \frac{1}{2} e^t = \cosh t,$$

so the cumulant generating function of  $X_1$  is

$$\Lambda(t) = \log \cosh t,$$

and indeed  $\Lambda(t) < \infty$  for all  $t \in \mathbb{R}$ . Then, as  $\frac{d}{dt}(tx - \log \cosh t) = x - \tanh t$ ,

$$\Lambda^*(x) = \sup_{t \in \mathbb{R}} (tx - \log \cosh t) = \operatorname{arctanh} x \cdot x - \log \cosh \operatorname{arctanh} x.$$

For  $x \in (-1, 1)$ ,

$$\operatorname{arctanh} x = \frac{1}{2} \log \frac{1+x}{1-x}.$$

Then

$$\cosh \operatorname{arctanh} x = \frac{1}{2} (e^{\operatorname{arctanh} x} + e^{-\operatorname{arctanh} x}) = \frac{1}{2} \sqrt{\frac{1+x}{1-x}} + \frac{1}{2} \sqrt{\frac{1-x}{1+x}} = \frac{1}{\sqrt{1-x^2}}.$$

With these identities,

$$\begin{aligned} \Lambda^*(t) &= \frac{x}{2} \log \frac{1+x}{1-x} + \frac{1}{2} \log(1-x^2) \\ &= \frac{x}{2} \log(1+x) - \frac{x}{2} \log(1-x) + \frac{1}{2} \log(1+x) + \frac{1}{2} \log(1-x) \\ &= \frac{1+x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x). \end{aligned}$$

With  $S_n = \sum_{k=1}^n X_k$ , applying Cramér's theorem, we get that for any  $a > E(X_1) = 0$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(S_n \geq an) = -\frac{1+x}{2} \log(1+x) - \frac{1-x}{2} \log(1-x).$$

For a Borel probability measure  $\mu$  on  $\mathbb{R}$ , we define its **Laplace transform**  $\check{\mu} : \mathbb{R} \rightarrow (0, \infty]$  by

$$\check{\mu}(t) = \int_{\mathbb{R}} e^{ty} d\mu(y).$$

Suppose that  $\int_{\mathbb{R}} |y| d\mu(y) < \infty$  and let  $M_1 = \int_{\mathbb{R}} y d\mu(y)$ , the first moment of  $\mu$ . For any  $t$  the function  $x \mapsto e^{tx}$  is convex, so by Jensen's inequality,

$$e^{tM_1} \leq \int_{\mathbb{R}} e^{ty} d\mu(y) = \check{\mu}(t).$$

Thus for all  $t \in \mathbb{R}$ ,

$$tM_1 - \log \check{\mu}(t) \leq 0.$$

For a Borel probability measure  $\mu$  with finite first moment, we define its **Cramér transform**  $I_{\mu} : \mathbb{R} \rightarrow [0, \infty]$  by<sup>7</sup>

$$I_{\mu}(x) = \sup_{t \in \mathbb{R}} (tx - \log \check{\mu}(t)).$$

For  $t = 0$ ,  $tx - \log \check{\mu}(t) = -\log \check{\mu}(0) = -\log(1) = 0$ , which shows that indeed  $0 \leq I_{\mu}(x) \leq \infty$  for all  $x \in \mathbb{R}$ . But  $tM_1 - \log \check{\mu}(t) \leq 0$  for all  $t$  yields

$$I_{\mu}(M_1) = 0.$$

---

<sup>7</sup>Heinz Bauer, *Probability Theory*, pp. 89–90, §12.