Mathematical Statistical Physics – LMU München, summer semester 2012
Hartmut Ruhl, Imre Péter Tóth
Homework sheet 2 – solutions – and exercises for the class on 27.04.2012

2.1 *Continuity of the measure*

(a) Prove the following:

Theorem 1 *(Continuity of the measure)*

i. If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $A_1, A_2, \dots$ is an increasing sequence of measurable sets (i.e. $A_i \in \mathcal{F}$ and $A_i \subset A_{i+1}$ for all i), then $\mu(\cup_{i=1}^{\infty} A_i) = \lim_{i\to\infty} \mu(A_i)$ (and both sides of the equation make sense).

ii. If $(\Omega, \mathcal{F}, \mu)$ is a measure space, $A_1, A_2, \dots$ is a decreasing sequence of measurable sets (i.e. $A_i \in \mathcal{F}$ and $A_i \supset A_{i+1}$ for all i) and $\mu(A_1) < \infty$, then $\mu(\cap_{i=1}^{\infty} A_i) = \lim_{i\to\infty} \mu(A_i)$ (and both sides of the equation make sense).

(b) Show that in the second statement the condition $\mu(A_1) < \infty$ is needed, by constructing a counterexample for the statement when this condition does not hold.

2.2 *Usefulness of the linearity of the expectation.* A building has 10 floors, not including the ground floor. On the ground floor, 10 people get into the elevator, and every one of them chooses a destination at random, uniformly out of the 10 floors, independently of the others. Let X denote the number of floors *on which the elevator stops* – i.e. the number of floors that were chosen by at least one person. Calculate the expectation and the variance of X. *(hint: First notice that the distribution of X is hard to calculate. Find a way to calculate the expectation and the variance without that.)*

2.3 **(homework)** Calculate the characteristic function of the normal distribution $\mathcal{N}(m, \sigma^2)$. (Remember the definition from the old times: $\mathcal{N}(m, \sigma^2)$ is the distribution on $\mathbb{R}$ with density (w.r.t. Lebesgues measure)

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

You can save yourself some paperwork if you only do the calculation for $\mathcal{N}(0,1)$ and reduce the general case to this using the relation between different normal distributions. You can and should use the fact that

$$\int_{-\infty}^{\infty} f_{m,\sigma^2}(x)\,\mathrm{d}x = 1$$

for every m and σ.

Solution: First we reduce the general case to the case of the standard normal distribution using the fact (known from old times, easy to check from the formulas) that if $X \sim \mathcal{N}(0,1)$ and $Y = m + \sigma X$, then $Y \sim \mathcal{N}(m, \sigma^2)$. As a result, the characteristic function for the normal distribution with expectation m and variance σ^2 is

$$\psi_{\mathcal{N}(m,\sigma^2)}(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{itm+it\sigma X}) = e^{itm}\mathbb{E}(e^{i(t\sigma)X}) = e^{itm}\psi_{\mathcal{N}(0,1)}(\sigma t), \tag{1}$$

where $\psi_{\mathcal{N}(0,1)}(t) := \mathbb{E}(e^{itX})$ is the characteristic function of the standard normal distribution.

Now we go on to calculate

$$\begin{aligned}\psi_{\mathcal{N}(0,1)}(t) &= \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2-2itx}{2}}\,\mathrm{d}x = \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-it)^2-(it)^2}{2}}\,\mathrm{d}x = e^{-\frac{t^2}{2}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-it)^2}{2}}\,\mathrm{d}x.\end{aligned}$$

We use the substitution $y := x - it$ to get

$$\psi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \,\mathrm{d}x = e^{-\frac{t^2}{2}}.$$

In the last step we used that the standard normal densitiy function (just like every probability density function) intergates to 1. Writing this back to (1), we get the final result

$$\psi_{\mathcal{N}(m,\sigma^2)}(t) = e^{itm} e^{-\frac{(\sigma t)^2}{2}}.$$

Remark: The substitution $y = x - it$ is not completely trivial to make rigorous. In fact, with this substitution, while x runs over the real line, y will run over a line in the complex plane, namely the line γ of complex numbers with imaginary part $-it$, so leaving the boundaries as $-\infty$ and ∞ after the substitution is cheating. To make the argument precise, one has to show that the integral on γ is equal to the integral on the real line. This is a typical application of a standard, but strong tool of complex analysis, called the *residue theorem.* I will not go into that here, and I don't expect the students to do so either.

2.4 *Dominated convergence and continuous differentiability of the characteristic function.*
The Lebesgue dominated convergence theorem is the following

Theorem 2 (dominated convergence) *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \dots$ measurable real valued functions on Ω which converge to the limit function pointwise, μ-almost everywehere. (That is, $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \Omega$, except possibly for a set of x-es with μ-measure zero.) Assume furthermore that the f_n admit a common* integrable dominating function*: there exists a $g : \Omega \to \mathbb{R}$ such that $|f_n(x)| \le g(x)$ for every $x \in \Omega$ and $n \in \mathbb{N}$, and $\int_\Omega g \,\mathrm{d}\mu < \infty$. Then (all the f_n and also f are integrable and)*

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu.$$

Use this theorem to prove the following

Theorem 3 (differentiability of the characteristic function) *Let X be a real valued random variable, $\psi(t) = \mathbb{E}(e^{itX})$ its characteristic function and $n \in \mathbb{N}$. If the n-th moment of X exists and is finite (i.e. $\mathbb{E}(|X|^n) < \infty$), then ψ is n times continuously differentiable and*

$$\psi^{(k)}(0) = i^k \mathbb{E}(X^k), \quad k = 0, 1, 2, \dots, n.$$

2.5 *Weak convergence and densities.*

(a) **(homework)** Prove the following

Theorem 4 *Let $\mu_1, \mu_2, \dots$ and μ be a sequence of probability distributions on $\mathbb{R}$ which are absolutely continouos w.r.t. Lebesgue measure. Denote their densities by $f_1, f_2, \dots$ and f, respectively. Suppose that $f_n(x) \overset{n\to\infty}{\longrightarrow} f(x)$ for every $x \in \mathbb{R}$. Then $\mu_n \Rightarrow \mu$ (weakly).*

(Hint: denote the cumulative distribution functions by $F_1, F_2, \dots$ and F, respectively. Use the Fatou lemma to show that $F(x) \le \liminf_{n\to\infty} F_n(x)$. For the other direction, consider $G(x) := 1 - F(x)$.

(b) Show examples of the following facts:

i. It can happen that the f_n converge pointwise to some f, but the sequence μ_n is not weakly convergent, because f is not a density.

ii. It can happen that the μ_n are absolutely continuous, $\mu_n \Rightarrow \mu$, but μ is not absolutely continuous.

iii. It can happen that the μ_n and also μ are absolutely continuous, $\mu_n \Rightarrow \mu$, but $f_n(x)$ does not converge to $f(x)$ for any x.

Solution:

(a) For every $n \in \mathbb{N}$ and $y \in \mathbb{R}$, the (cumulative) distribution function F_n of μ_n at y can be calculated as

$$F_n(y) = \int_{(-\infty,y]} f_n(x)\,\mathrm{d}x,$$

while the distribution function of μ at y is

$$F(y) = \int_{(-\infty,y]} f(x)\,\mathrm{d}x.$$

Since by assumption $f(x) = \lim_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)$, the Fatou lemma implies

$$F(y) = \int_{(-\infty,y]} \liminf_{n\to\infty} f_n(x)\,\mathrm{d}x \le \liminf_{n\to\infty} \int_{(-\infty,y]} f_n(x)\,\mathrm{d}x = \liminf_{n\to\infty} F_n(x). \tag{2}$$

A completely similar argument works for the "tail distribution functions"

$$G_n(y) = 1 - F_n(y) = \int_{[y,\infty)} f_n(x)\,\mathrm{d}x, \quad G(y) = 1 - F(y) = \int_{[y,\infty)} f(x)\,\mathrm{d}x,$$

which gives

$$G(y) \le \liminf_{n\to\infty} G_n(x).$$

But this implies

$$F(y) = 1 - G(y) \ge 1 - \liminf_{n\to\infty} G_n(x) = \limsup_{n\to\infty}(1 - G_n(x)) = \limsup_{n\to\infty} F_n(x).$$

This, together with (2) implies

$$F(y) = \lim_{n\to\infty} F_n(y) \qquad \text{for every } y \in \mathbb{R},$$

which is known from the lecture to be equivalent to weak convergence of the measures.

2.6 For a γ-detector, the times $\tau_1, \tau_2, \ldots$ that elapse between consecutive hits are independent random varibales which are exponentially distributed with $\mathbb{E}\tau_i = 1$. That is, their common density is $f(x) = e^{-x}\mathbf{1}_{[0,\infty)}(x)$ (where $\mathbf{1}$ stands for indicator function). (We measure time in seconds.)

Use the Cramer large deviation theorem to estimate the probability that we have to wait less than 500 seconds for the 1000-th hit.

2.7 **(homework)** Let $X_i, X_2, \ldots, X_n$ be independent random variables with Poisson distribution $X_i \sim Poi(\lambda)$. (That is, $\mathbb{P}(X_i = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$.) Use the Cramer large deviation theorem to estimate the probability $\mathbb{P}(\sum_{k=1}^{n} X_k > 1000)$

(a) for $\lambda = 1$, $n = 500$,

(b) for $\lambda = 500$, $n = 1$.

Solution: To apply the Cramer theorem, we first need to calculate the rate function for the Poisson distribution. The first step in the calculation is the moment generating function Z. Unlike in the lecture, I will denote the argument of Z by t (and not λ, because λ is already used for the parameter of the Poisson distribution). So

$$Z(t) := \mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \lambda)^k}{k!} = e^{-\lambda} e^{e^t \lambda} = e^{\lambda(e^t - 1)}.$$

(This exists (the sum converges) for every $t \in \mathbb{R}$.) The second step is the logarithmic moment generating function:

$$\hat{I}(t) := \log Z(t) = \lambda(e^t - 1).$$

The rate function $I(x)$ will be obtained as the Legendre transform of $\hat{I}$ defined as $I(x) := \sup_{t \in \mathbb{R}} \{xt - \hat{I(t)}\}$. This is not difficult, since for every x which is of interest (see later), the supremum is known to be a maximum, which is obtained at exactly one $t^* = t^*_x$ given by $x = \hat{I}'(t^*)$. In our case

$$x = \hat{I}'(t^*) = \lambda e^{t^*}, \quad \text{so} \quad t^* = \log \frac{x}{\lambda},$$

and the supremum is

$$I(x) = xt^* - \hat{I}(t^*) = x \log \frac{x}{\lambda} - \lambda(\frac{x}{\lambda} - 1) = x \log \frac{x}{\lambda} - x + \lambda.$$

Note that this calculation works for exactly those values of x, which are obtained as values of $\hat{I}'$, in our case $x \in (\underline{x}, \overline{x}) = (0, \infty)$. This is exactly the (open) interval in which the average of a Poissonian sample can be. So to be absolutely precise,

$$I(x) = \begin{cases} x \log \frac{x}{\lambda} - x + \lambda & \text{if } x \in (0, \infty), \\ 0 & \text{if not.} \end{cases}$$

Now we are ready to apply the Cramer theorem to estimate the desired probability

$$\mathbb{P}(\sum_{k=1}^{n} X_k > 1000).$$

We will use the notation $S_n = \sum_{k=1}^{n} X_k$.

(a) for $\lambda = 1$, $n = 500$:

$$\mathbb{P}(S_n > 1000) = \mathbb{P}(\frac{S_n}{n} \in (2, \infty)) \approx e^{-n \inf_{x \in (2,\infty)} I(x)} = e^{-500 I(2)} \approx e^{-193.15} \approx 1.31 \cdot 10^{-84}.$$

(b) for $\lambda = 500$, $n = 1$:

$$\mathbb{P}(S_n > 1000) = \mathbb{P}(\frac{S_n}{n} \in (1000, \infty)) \approx e^{-n \inf_{x \in (1000,\infty)} I(x)} = e^{-1 I(1000)} \approx e^{-193.15} \approx 1.31 \cdot 10^{-84}.$$

Remark: It is no surprise that the two results are equal. Indeed, the two S_n of the two settings are identically distributed – both are $Poi(500)$, since the sum of independent Poissonians is also Poissonian, with the parameters (= expected values) added. This also shows that when applying the Cramer theorem for a Poisson distribution, n being large or not plays no role. On the other hand, as example (b) shows, already for $n = 1$ the theorem says something about the tail of the distribution function, which is not completely immediate from summation of the discrete distribution.

It is also no surprise that the result is the same as in the previous exercise about exponentially distributed random variables. Indeed, with the notation of the two exercises,

$$\{\tau_1 + \tau_2 + \tau_{1000} \leq 500\} = \{S_n \geq 1000\}$$

– i.e. the two events are the same –, if S_n denotes the number of hits during the first 500 seconds, which is known to be distributed as $Poi(500)$. This intimate relationship between the exponential and the Poisson distribution is the key to the mathematical model of the hitting times, called a Poisson process.

2.8 *Change of measure in the proof of the Cramer theorem.* Let μ be a probability distribution on $\mathbb{R}$ and $Z(\lambda) := \int_{\mathbb{R}} e^{\lambda x}\,\mathrm{d}\mu(x)$ its moment generating function. Suppose that $Z(\lambda)$ is finite on the interval $(\underline{\lambda}, \overline{\lambda})$ with $\underline{\lambda} < 0 < \overline{\lambda}$. Let $\hat{I}(\lambda) := \log Z(\lambda)$, $y \in \mathbb{R}$ and suppose that $\lambda^* \in (\underline{\lambda}, \overline{\lambda})$ can be chosen such that $\hat{I}'(\lambda^*) = y$. Now let μ^* be the probability distribution on $\mathbb{R}$ which is absolutely continuous w.r.t. μ, and its density is $\frac{1}{Z(\lambda^*)} e^{\lambda^* x}$ – that is,

$$\mathrm{d}\mu^*(x) = \frac{1}{Z(\lambda^*)} e^{\lambda^* x}\,\mathrm{d}\mu(x).$$

(a) **(homework)** Show that the expectation of μ^* is exactly y – that is, $\int_{\mathbb{R}} x\,\mathrm{d}\mu^* = y$. (Don't worry much about exchanging differentiation and integrals.)

(b) Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with distribution μ, and let $X_1^*, X_2^*, \ldots, X_n^*$ be i.i.d random variables with distribution μ^*. Denote the distribution of $S_n := X_1 + X_2 + \cdots + X_n$ by μ_n and the distribution of $S_n^* := X_1^* + X_2^* + \cdots + X_n^*$ by μ_n^*. Show that

$$\mathrm{d}\mu_n^*(x) = \frac{1}{Z(\lambda^*)^n} e^{\lambda^* x}\,\mathrm{d}\mu_n(x).$$

(Hint: condsider the *joint distribution* of $(X_1^*, X_2^*, \ldots, X_n^*)$ (on $\mathbb{R}^n$). How is this related to the joint distribution of $(X_1, X_2, \ldots, X_n)$?)

Solution:

(a) Formally differentiating $Z(\lambda) := \int_{\mathbb{R}} e^{\lambda x}\,\mathrm{d}\mu(x)$, we get

$$Z'(\lambda) := \int_{\mathbb{R}} x e^{\lambda x}\,\mathrm{d}\mu(x).$$

It can be shown – similarly to Exercise 4 – that this is indeed true for $\lambda \in (\underline{\lambda}, \overline{\lambda})$, and this is what you didn't have to worry about. Having that, we use the definition of μ^* to calculate

$$\int_{\mathbb{R}} x\,\mathrm{d}\mu^*(x) = \int_{\mathbb{R}} x \frac{1}{Z(\lambda^*)} e^{\lambda^* x}\,\mathrm{d}\mu(x) = \frac{\int_{\mathbb{R}} x e^{\lambda^* x}\,\mathrm{d}\mu(x)}{Z(\lambda^*)} = \frac{Z'(\lambda^*)}{Z(\lambda^*)}.$$

Now by the definition of $\hat{I}$ and λ^*, this is further equal to

$$\int_{\mathbb{R}} x\,\mathrm{d}\mu^*(x) = \frac{Z'(\lambda^*)}{Z(\lambda^*)} = \hat{I}'(\lambda^*) = y.$$