Mathematical Statistical Physics – LMU München, summer semester 2012

Hartmut Ruhl, Imre Péter Tóth

Homework sheet 4 – due on 18.05.2012 – and exercises for the class on 11.05.2012

4.1 Γ *function and polar coordinates practice.* Calculate the ($(n-1)$-dimensional) surface volume $s_n(r)$ of the n-dimensional sphere with radius r, for every positive integer n in terms of the Γ function defined as

$$\Gamma(x) := \int_0^\infty t^{x-1} e^{-t} \,\mathrm{d}t.$$

hint: integrate $f(x_1,\dots,x_n) = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{x_1^2+\dots+x_n^2}{2}}$ *on* $\mathbb{R}^n$.

4.2 Let the random vector $(v_1, v_2, \dots, v_N)$ be uniformly distributed on the (surface of the) N-dimensional sphere with radius $\sqrt{2NE}$, where $E \in (0,\infty)$ is a fixed number. Find the limit distribution of v_1 as $N \to \infty$. *hint: calculate the density for each N using the result of Exercise 1, then use the Stirling formula*

$$\Gamma(x) = \sqrt{\frac{2\pi}{x}} \left(\frac{x}{e}\right)^x (1 + o(1)).$$

4.3 **(homework)** Let the random vector $(v_1, v_2, \dots, v_N)$ be uniformly distributed on the simplex

$$\{(v_1,\dots,v_N) \in \mathbb{R}^N \,:\, 0 \le v_i,\, v_1 + \dots + v_N = NE\},$$

where $E \in (0,\infty)$ is a fixed number. Find the limit distribution of v_1 as $N \to \infty$.

Solution 1: explicit calculation. Let $A_n(r)$ denote the $(n-1)$-dimensional surface volume of the simplex $\{(x_1,\dots,x_n) \in \mathbb{R}^n \,:\, 0 \le x_i,\, x_1 + \dots + x_n = r\} \subset \mathbb{R}^n$. As a warming-up, we consider how we could calculate the function A_n if we knew A_{n-1}. The r-dependence is trivial from the scaling of volume with linear size: $A_n(r) = C_n r^{n-1}$. To calculate $C_n = A_n(1)$, we integrate

$$C_n = \int_0^1 A_{n-1}(r(x_1)) c_n \,\mathrm{d}x_1,$$

where $r(x)$ is the sum $x_2 + \dots + x_n$ under the condition $x_1 = x$, so simply $r(x_1) = 1 - x_1$. A bit more interesting is the number c_n, which has the meaning that the two submanifolds $\{x_1 = x\}$ and $\{x_1 = x + \mathrm{d}x\}$ have the distance $c_n \,\mathrm{d}x$. This c_n is indeed a constant (not a function of x), since our surface is flat. It wouldn't be hard to find out the value from a geometrical consideration, but we don't really need it, so we just go on with the notation, and get

$$C_n = \int_0^1 C_{n-1}(1 - x_1)^{n-2} c_n \,\mathrm{d}x_1.$$

Having that considered, we return to the original problem. The density of v_1 (with some fixed N) is

$$f_N(x) = \frac{C_{n-1}(NE - x)^{n-2} c_N}{A_N(NE)}, \quad \text{for } 0 \le x \le NE.$$

We don't (need to) know the values of C_{N-1}, c_N and C_N, but $K_{N,E} := \frac{C_{N-1} c_N}{A_N(NE)}$ has to be the appropriate normalizing factor so that $f_N(x)$ is indeed a density, so

$$1 = \int_{-\infty}^\infty f_N(x) \,\mathrm{d}x = K_{N,E} \int_0^{NE} (NE - x)^{N-2},$$

which leads to $K_{N,E} = \frac{N-1}{(NE)^{N-1}}$ and

$$f_N(x) = \begin{cases} \frac{N-1}{(NE)^{N-1}}(NE-x)^{N-2}, & \text{if } 0 \le x \le NE, \\ 0, & \text{if not.} \end{cases}$$

With $N \to \infty$, this is easily seen to converge pointwise to

$$f_(x) = \lim_{N\to\infty} f_N(x) = \begin{cases} \frac{1}{E}e^{-\frac{x}{E}}, & \text{if } 0 \le x, \\ 0, & \text{if not.} \end{cases}$$

This is the density of the exponential distribution with parameter $\frac{1}{E}$, so by the statement of Exercise 2.5(a), v_1 converges weakly to $Exp(\frac{1}{E})$.

(Note that a reference to Exercise 2.5(a) is not really important here: the cumulative distribution function $F_N(x) := \int_{-\infty}^{x} f_N(y)\,\mathrm{d}y$ can be calculated explicitely, and its pointwise convergence to $F(x) = (1 - e^{-\frac{x}{E}})\mathbf{1}_{[0,\infty)}(x)$ can be checked directly.)

Solution 2: making use of the friendship between multiplication, addition and the exponential function. Let X_1,X_2,…,X_N be i.i.d. random variables distributed as $Exp(\lambda)$, with any λ. Their joint density $F(x_1, x_2, \dots, x_N) = \lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}\dots\lambda e^{-\lambda x_N}$ is constant on the simplex $\{x_1 + x_2 + \dots + x_N = NE,\ x_i \ge 0\}$, so the conditional distribution of X_1,X_2,…,X_N under the condition $X_1 + X_2 + \dots + X_N = NE$ is exactly the uniform measure on the simplex. That means, the distribution of v_1 we are looking for is exactly the same as the conditional distribution of X_1 under the condition $X_1 + X_2 + \dots + X_N = NE$. To calculate this, we will use some knowledge of probability, which is elementary, but was not part of this course.

Introduce the notation $U := X_1$, $V := X_1+X_2+\dots+X_N$. Now U is distributed as $Exp(\lambda)$, and the distribution of V is also well knonw: it's called the gamma distribution with parameters (N, λ) and has the density

$$f_V(v) = \begin{cases} \frac{\lambda^N}{\Gamma(N)}v^{N-1}e^{-\lambda v}, & \text{if } v \ge 0, \\ 0, & \text{if not.} \end{cases}$$

We also need the *joint density* of (U, V). For this purpose, we introduce $X := X_1$ and $Y := X_2 + \dots + X_N$. The joint distribution of (X, Y) is easy, because the are independent (unlike (U, V)), $X \sim Exp(\lambda)$ and $Y \sim Gamma(N-1, \lambda)$:

$$f_{X,Y}(x,y) = \begin{cases} \lambda e^{-\lambda x}\frac{\lambda^{N-1}}{\Gamma(N-1)}y^{N-2}e^{-\lambda y}, & \text{if } x \ge 0 \text{ and } y \ge 0, \\ 0, & \text{if not.} \end{cases}$$

We obtain (U, V) as a (linear) transformation of (X, Y):

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}.$$

The matrix $J = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is also the Jacobian of the mapping $(X, Y) \to (U, V)$, so the densitiy transformation rule gives (with the notation $u = x$, $v = x + y$)

$$f_{U,V}(u,v) = \frac{1}{|\det(J)|}f_{X,Y}(x,y) = \frac{1}{1}f_{X,Y}(u, v-u) = \begin{cases} \frac{\lambda^N}{\Gamma(N-1)}(v-u)^{N-2}e^{-\lambda v}, & \text{if } v \ge u \ge 0, \\ 0, & \text{if not.} \end{cases}$$

The conditional density we are looking for is

$$f_{v_1}(u) = f_{U|V}(u|V = NE) = \frac{f_{U,V}(u, NE)}{f_V(NE)} = \begin{cases} (N-1)\frac{(NE-u)^{N-2}}{(NE)^{N-1}}, & \text{if } 0 \le u \le NE, \\ 0, & \text{if not.} \end{cases}$$

Now we can check the pointwise convergece of the density, or (if you like) calculate the distribution function and check the pointwise convergence of that, as in the first solution. Anyway we get $v_1 \Rightarrow Exp(\frac{1}{E})$.

Solution 3: heuristically, if we know the result in advance. Let X_1,X_2,...,X_N be i.i.d. random variables distributed as $Exp(\lambda)$ as in the previous solution, but now we set $\lambda = \frac{1}{E}$. Now the distribution of v_1 we are lookiong for is exactly the conditional distribution of X_1 under the condition $X_1 + X_2 + \cdots + X_N = NE$ (see the previous solution for the argument). But now the condition is exactly that

$$\frac{X_1 + X_2 + \cdots + X_N}{N} = \mathbb{E}X_1,$$

and the law of large numbers states that this always happens – at least in some asymptotic sense, when $N \to \infty$. so for $N \to \infty$ this condition is empty (meaning a set of probability 1). So the conditional distribution is the same as the unconditional distribution, so $v_1 \Rightarrow X_1 \sim Exp(\frac{1}{E})$. This argument can be made precise by allowing some ε deviation from the mean in the condition, and being careful enough when exchanging the limits $N \to \infty$ and $\varepsilon \to 0$.

4.4 We roll a fair die 10 times and record the results. Let X be the random 10-digit number we get. Calculate the entropy of X.

4.5 **(homework)** We toss a biased coin with $\mathbb{P}(heads) = p \in (0, 1)$ 10 times and record the results. Let Y be the random 10-long string we get. Calculate the entropy of Y.

Solution 1: brute force calculation. Use the notation $q = 1 - p$. The experiment has 2^{10} possible outcomes, out of which $\binom{10}{k}$ consist of k heads and $n - k$ tails (in some order, for every $k \in \{0, 1, \ldots, 10\}$). These $\binom{10}{k}$ outcomes have probability $p^k q^{10-k}$). So the definition of entropy gives

$$S = -\sum_{i=1}^{2^{10}} p_i \log p_i = -\sum_{k=0}^{10} \binom{10}{k} p^k q^{10-k} \log(p^k q^{10-k}).$$

We use $\log(p^k q^{10-k}) = 10 \log q + k(\log p - \log q)$ to get

$$S = -\left[(10 \log q) \sum_{k=0}^{10} \binom{10}{k} p^k q^{10-k} + (\log p - \log q) \sum_{k=0}^{10} k \binom{10}{k} p^k q^{10-k} \right].$$

The coefficient of $(10 \log q)$ is the sum of the probabilities in a binomial distribution with parameters $(10, p)$, so it is $1 = p + q$. The coefficient of $(\log p - \log q)$ is exactly the expectation of this binomial distribution, so it is $10p$ (see also exercise 1.1(b)). We got

$$S = -\left[(10 \log q)(p + q) + (\log p - \log q)10p \right] = -10(p \log p + q \log q).$$

Solution 2: additivity of the entropy. Use the notation $q = 1 - p$. The entropy of a sequence of *indepentent* random variables is the sum of the entropies, so the entropy we are looking for is 10 times the entropy of the outcome of a single coint toss. A single toss has two possible outcomes with probabilities p and q, so its entropy is by definition $-(p\log p + q\log q)$. The entropy of the sequence is thus

$$S = -10(p\log p + q\log q).$$

4.6 *Maximum entropy principle.* The maximum entropy principle describes the probability measures that have maximum relative entropy w.r.t some reference measure under certain constraints – namely, with the integrals of certain (arbitrary) functions being pre-given:

Theorem 1 (Maximum entropy principle) *Let $(\Omega, \mathcal{F}, \nu)$ be a (not necessarily probability) measure space. Suppose that $X_1, \ldots, X_n$ are pre-given measurable (real-valued) functions on $(\Omega, \mathcal{F})$ and $m_1, \ldots, m_n$ are pre-given real numbers. We consider those probability measures on $(\Omega, \mathcal{F})$, w.r.t. which the integrals of our pre-given functions are exactly the pre-given numbers:*

$$\mathcal{P}(\underline{X}, \underline{m}) := \left\{ \mu \text{ probability measure on } (\Omega, \mathcal{F}) \, : \, \int_\Omega X_i \, \mathrm{d}\mu = m_i \text{ for } i = 1, \ldots, n \right\}.$$

Suppose that we can choose $t_1, \ldots, t_n \in \mathbb{R}$ with the following properties:

- $Z_{\underline{t}} := \int_\Omega e^{-\sum_{i=1}^n t_i X_i(\omega)} \, \mathrm{d}\nu(\omega) < \infty$,
- *the probability measure $\mu_{\underline{t}}$ on $(\Omega, \mathcal{F})$ which is absolutely continuous w.r.t. ν, with density $\rho_{\underline{t}}(\omega) := \frac{1}{Z_{\underline{t}}} e^{-\sum_{i=1}^n t_i X_i(\omega)}$ satisfies $\mu_{\underline{t}} \in \mathcal{P}(\underline{X}, \underline{m})$. Then $\mu_{\underline{t}}$ is the (unique) probability measure in $\mathcal{P}(\underline{X}, \underline{m})$ which has maximal entropy w.r.t ν, and*

$$S(\mu_{\underline{t}}; \nu) = \sum_{i=1}^n t_i m_i + \log Z_{\underline{t}}.$$

Use this theorem to find (the distribution of) the random variable X with maximum entropy (if it exist)

(a) w.r.t. Lebesgue measure on $\mathbb{R}$, under the constraint $\mathbb{E}X = m$,

(b) w.r.t. Lebesgue measure on $\mathbb{R}^+$, under the constraint $\mathbb{E}X = m$,

(c) **(homework)** w.r.t. Lebesgue measure on $\mathbb{R}$, under the constraints $\mathbb{E}X = m$, $\mathrm{Var}X = v$,

(d) **(homework)** w.r.t. the counting measure on $\mathbb{N}$, under the constraint $\mathbb{E}X = m$.

Solution:

(a) Doesn't exist (discussed in class).

(b) Exponential distribution with ecpectation m (or parameter $\frac{1}{m}$) (discussed in class).

(c) For $v < 0$ the exercise makes no sense. For $v = 0$ the only probability distribution satisfying the constraints is the Dirac measure concentrated on m, so this is also the distribution with maximum entropy (although the entropy is $-\infty$). From now on we suppose $v > 0$.

Use the maximum entropy principle with $(\Omega, \mathcal{F}, \nu) = (\mathbb{R}, \mathcal{B}, \mathrm{Leb})$, $n = 2$, $X_1(x) = x$, $m_1 = m$, $X_2(x) = x^2$ and $m_2 = v + m^2$. Then the two constaints ensure exactly that the expectation is m and the variance is v. The theorem ensures that if a probability density of the form $f(x) = \mathit{const}\, e^{-(t_1 x + t_2 x^2)}$ exists with expectation m and second moment $v + m^2$, then it is the density of the unique probability distribution with maximum entropy satisfying the constraints. But yes, of course, a Gaussian density is exactly of this form, and will satisfy the contraints exactly if it has parameters m and v, so

$$f(x) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-m)^2}{2v}}$$

will do.

(d) First notice that the *density w.r.t. the counting measure* is nothing else than the discrete probability distribution for discrete random variables. The solution depends slightly on whether we mean $N = \{1, 2, \dots\}$ or $N = \{0, 1, 2, \dots\}$, but don't worry about that first. Use the maximum entropy principle with $(\Omega, \mathcal{F}, \nu) = (\mathbb{N}, 2^{\mathbb{N}}, \chi)$ (with χ denoting the counting measure), $n = 1$, $X_1(n) = n$ and $m_1 = m$. Then the constaint ensures exactly that the expectation is m. The theorem ensures that if a probability sequence of the form $p_n = const\, e^{-t_1 n}$ exists with expectation m, then it is the unique discrete probability distribution with maximum entropy satisfying the constraint. But yes, of course, the geometrical distribution is exactly of this form, and will satisfy the constraint if we choose the parameter properly:

- With the convention $N = \{1, 2, \dots\}$, we set $p_n = (1-p)p^{n-1}$ for $n = 1, 2, \dots$. This leads to the expectation being $\frac{1}{p}$, so we have to choose $p = \frac{1}{m}$.
- With the convention $N = \{0, 1, 2, \dots\}$, we set $p_n = (1-p)p^n$ for $n = 0, 1, 2, \dots$. This leads to the expectation being $\frac{1}{p} - 1$, so we have to choose $p = \frac{1}{m+1}$.

Note that the question makes no sense if $m < 1$ (or $m < 0$, depending on the convention on $\mathbb{N}$). For e.g. $m = 1$ and $\mathbb{N} = \{1, 2, \dots\}$, the only prob. distribution satisfying the constraint is the Dirac measure concentrated on 1, so it is also the prob. distribution with maximum entropy (although the entropy is 0).

Remark: The above geometrical distribution is often called *pessimistic* for $N = \{1, 2, \dots\}$ and *optimistic* for $N = \{0, 1, 2, \dots\}$. Guess why.

4.7 *Microcanonical description of the free gas.* Consider N identical particles of mass m in a box $\Lambda \subset \mathbb{R}^3$ (with volume V), with the Hamiltonian

$$H(\underline{q}, \underline{p}) = \sum_{i=1}^{N} \frac{\vec{p_i}^2}{2m}$$

(the particles are non-interacting). Fix the total energy to be E.

(a) Describe the microcanonical distribution $\mu_{\text{micr}} = \mu_{N,V,E}$.

(b) Calculate the microcanonical partition function $Z_{\text{micr}} = Z(N, V, E)$. (Use the result of Exercise 1.)

(c) Calculate the entropy $S(N, V, E)$ of μ_{micr} (relative to the "natural reference measure", which is the conditional measure of the Lebesgue measure (of the phase space) on the $\{H = E\}$ surface).

(d) Set $E = Nu$, $V = Nv$ with u, v fixed constants, so $S(N, V, E)$ becomes $S_{u,v}(N)$. How does $S_{u,v}(N)$ scale with N? Use the Stirling formulas

$$\Gamma(x) = \sqrt{\frac{2\pi}{x}} \left(\frac{x}{e}\right)^x (1 + o(1)) \quad , \quad n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n (1 + o(1)).$$

Have you not forgotten to factorize the phase space due to the indistinguishability of the particles?

Solution:

(a) Let $\sim$ denote the equivalence relation on Λ^N which identifies sequences that can be obtained from each other by permutation. The microcanonical distribution is the uniform distribution on $\tilde{\Lambda}_N = \Lambda^N/\sim$ times the uniform distribution on the moment sphere

$$S_{3N}(\sqrt{2mE}) = \left\{ (\vec{p_1}, \vec{p_2}, \dots, \vec{p_N}) \, : \, \sum_{i=1}^{N} \vec{p_i}^2 = 2mE \right\} \subset \mathbb{R}^{3N}.$$

For that we didn't need to know the "volume" of this set w.r.t. the reference measure: it cancels out anyway with the normalization.

(b) The partition function $Z_{\text{micr}} = Z(N, V, E)$ is the volume of $\tilde{\Lambda}_N \times S_{3N}(\sqrt{2mE})$ (which is the $\{H = E\}$ surface) w.r.t. the reference measure, which is the (non-normalized) restriction of the Liouville measure to this set. We know that this reference measure has density $\frac{1}{|\nabla H|}$ w.r.t. $\text{Leb}_{\tilde{\Lambda}_N} \otimes \text{Leb}_{S_{3N}(\sqrt{2mE})}$, where $\text{Leb}_{S_{3N}(\sqrt{2mE})}$ is just a notation for the surface measure on the sphere. The gradient in the formula for the density has zero configurational component, and the velocity (more precisely, moment) component is radial, with length $\frac{1}{2m}2r$ if r denotes the distance from the origin (because $\frac{\mathrm{d}}{\mathrm{d}r}r^2 = 2r$). So

$$\frac{1}{|\nabla H|} = \frac{2m}{2\sqrt{\sum_i \vec{p_i}^2}} = \frac{2m}{2\sqrt{2mH}} = \sqrt{\frac{m}{2H}}.$$

We happily see that the density is constant on the $\{H = E\}$ set and the value is $\sqrt{\frac{m}{2E}}$. (The answer to part (a) is actually only verified now.) Now we can calculate

$$\begin{aligned} Z(N, V, E) &= \int_{\tilde{\Lambda}_N \times S_{3N}(\sqrt{2mE})} \sqrt{\frac{m}{2E}} \, \mathrm{d}(\text{Leb}_{\tilde{\Lambda}_N} \otimes \text{Leb}_{S_{3N}(\sqrt{2mE})}) \\ &= \sqrt{\frac{m}{2E}} \text{Leb}_{\tilde{\Lambda}_N}(\tilde{\Lambda}_N) \text{Leb}_{S_{3N}(\sqrt{2mE})}(S_{3N}(\sqrt{2mE})) \\ &= \sqrt{\frac{m}{2E}} \frac{\text{Leb}(\Lambda)^N}{N!} s_{3N}(\sqrt{2mE}) = \sqrt{\frac{m}{2E}} \frac{V^N}{N!} s_{3N}(\sqrt{2mE}) \end{aligned}$$

with $s_n(r) = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} r^{n-1}$ from Exercise 1. Putting this together, we get

$$Z(N, V, E) = \sqrt{\frac{m}{2E}} \frac{V^N}{N!} \frac{2\pi^{3N/2}}{\Gamma(\frac{3N}{2})} (2mE)^{\frac{3N-1}{2}} = \frac{V^N}{EN!\Gamma(\frac{3N}{2})} (2\pi mE)^{\frac{3N}{2}}.$$

(c) The entropy of the uniform distribution (w.r.t. the reference measure) is alway the logarithm of the volume:

$$S = -\int \frac{1}{Z} \log \frac{1}{Z} \, \mathrm{d}\mu_{ref} = -\frac{1}{Z} log \frac{1}{Z} \cdot Z = \log Z,$$

so

$$S(N, V, E) = \log Z(N, V, E) = N \log V + \frac{3N}{2} \log(2\pi mE) - \log(EN!\Gamma(\frac{3N}{2})).$$

(d) Setting $E = Nu$ and $V = Nv$ we get

$$S_{u,v}(N) = N \log v + \frac{3N}{2} \log(2\pi m u) - \log u + \log \frac{N^N N^{\frac{3N}{2}}}{N N! \Gamma(\frac{3N}{2})}$$

In the argument of the last logarithm we use the Stirling formulas to get

$$\frac{N^N N^{\frac{3N}{2}}}{N N! \Gamma(\frac{3N}{2})} = \frac{N^N N^{\frac{3N}{2}}}{N\sqrt{2\pi N}\left(\frac{N}{e}\right)^N \sqrt{\frac{4\pi}{3N}}\left(\frac{3N}{2e}\right)^{\frac{3N}{2}}}(1+o(1)) = \sqrt{\frac{3}{2}\frac{1}{2\pi}}\frac{1}{N}e^N\left(\frac{2e}{3}\right)^{\frac{3N}{2}}(1+o(1)).$$

Writing this back,

$$S_{u,v}(N) = N\left[\log\left(v(mu)^{3/2}\right) + \frac{3}{2}\log\frac{4\pi}{3} + \frac{5}{2}\right] - \log N + \left[\log\left(\sqrt{\frac{3}{2}\frac{1}{2\pi}}\right) - \log u\right] + o(1).$$

The essence of this is that

$$\frac{S_{u,v}(N)}{N} \xrightarrow{N\to\infty} \log\left(v(mu)^{3/2}\right) + \frac{3}{2}\log\frac{4\pi}{3} + \frac{5}{2}.$$

If I had forgotten to factorize the phase space due to the indistinguishability of the particles, then the partition sum $Z(N, V, E)$ would have an extra $N!$ factor. As a consequence, the leading term in $S_{u,v}(N)$ would be of order $N \log N$ coming from this factorial, and $\frac{S_{u,v}(N)}{N}$ would not converge.
(Remark: The argument $v(mu)^{3/2}$ of the logarithm in the limiting entropy is not unitless, so the logarithm depends on the choice of units. When calculating thermodynamic quantities as derivatives of the entropy, this uncertainty only effects the value of the chemical potential, up to an additive constant. So the "Physics of the system" is not affected.)