Limit/large dev. thms. HW assignment 7, SOLUTIONS

1. (15 marks) Decide about the following functions $\mathbb{R} \to \mathbb{C}$ whether they are characteristic functions of probability distributions or not.

$$(a)\ \frac{1}{1+t^2}, \quad (b)\ \exp(-t^4), \quad (c)\ \sin(t), \quad (d)\ \cos(t), \quad (e)\ \frac{1+\cos t}{2}, \quad (f)\ \frac{\sin(t)}{t} \quad (g)\ 2\frac{1-\cos(t)}{t^2}$$

Hint: You will NOT need to use Bochner's theorem (page 88). To show that a function is not a characteristic function, you need to use the properties that we have learnt on the April 7 lecture. To show that it is a characteristic function of a random variable X, you have to find the distribution of X. It is a wise idea to read the official solution of HW6.2 before you start solving this exercise.

Solution:

(a) This was solved in HW6.2(a): if $f(x) = \frac{1}{2}e^{-|x|}$ then $\int_{-\infty}^{\infty} e^{itx} f(x)\,\mathrm{d}x = \frac{1}{1+t^2}$.

(b) $\exp(-t^4)$ is not a characteristic function, because if we assume that $\exp(-t^4) = \varphi(t) = \mathbb{E}(e^{itX})$ for some X, then $\mathbb{E}(X^2) = -\varphi''(0) = 0$, therefore $X = 0$, but $\mathbb{E}(e^{it0}) = 1 \neq \exp(-t^4)$, a contradiction.

Remark: This solution had a little gap in it. Namely, in the theorem on page 90 of the scanned lecture notes, we only proved $\mathbb{E}(X^2) = -\varphi''(0)$ under the assumption $E(X^2) < +\infty$. In our current case nothing guarantees that this theorem can be applied. In a later homework we will return to this problem and fix the gap.

(c) $\sin(0) = 0 \neq 1$, so $\sin(t)$ cannot be a characteristic function.

(d) If $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$, then $\mathbb{E}(e^{itX}) = \frac{1}{2}e^{it\cdot 1} + \frac{1}{2}e^{it\cdot(-1)} = \cos(t)$.

(e) If $\mathbb{P}(X = 0) = \frac{1}{2}$ and $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{4}$, then $\mathbb{E}(e^{itX}) = \frac{1}{4}e^{it\cdot 1} + \frac{1}{2}e^{it\cdot 0} + \frac{1}{4}e^{it\cdot(-1)} = \frac{1+\cos t}{2}$

(f) We have seen in class (see page 89 of scanned) that if $X \sim \mathrm{UNI}[-1, 1]$ then $\mathbb{E}(e^{itX}) = \frac{\sin(t)}{t}$.

(g) We have seen in HW6.2(b) that $\left(\frac{\sin(t/2)}{t/2}\right)^2$ is the characteristic function of the sum of two independent random variables with $\mathrm{UNI}[-\frac{1}{2}, \frac{1}{2}]$ distribution. Now

$$\left(\frac{\sin(t/2)}{t/2}\right)^2 = 4\frac{\sin^2(t/2)}{t^2} = 4\frac{(1-\cos(t))/2}{t^2} = 2\frac{1-\cos(t)}{t^2}$$

2. (10 marks) In this exercise $\sqrt{z}$ denotes the complex analytic function which is defined for all complex numbers except for the negative real numbers in the following way: if $\mathrm{Im}(z) \geq 0$ then $\arg(\sqrt{z}) = \frac{1}{2}\arg(z)$ and $|\sqrt{z}| = \sqrt{|z|}$, moreover we extend the function to the half-plane $\mathrm{Im}(z) \leq 0$ by the identity $\sqrt{\overline{z}} = \overline{\sqrt{z}}$. Then of course $z \geq 0$ implies $\sqrt{z} \geq 0$, so this complex function $\sqrt{\cdot} : \mathbb{C} \setminus \{-\mathbb{R}_+\} \to \mathbb{C}$ is an analytic extension of the usual square root function $\sqrt{\cdot} : \mathbb{R}_+ \to \mathbb{R}_+$

We consider simple symmetric random walk on $\mathbb{Z}$. Recall that we denote by T_k the time it takes to reach level k (see page 59 of the scanned lecture notes). Recall that we denote by R_k the time of the k'th return to the origin (see page 65). We have found the generating function of T_1 in class (page 99).

(a) Find the generating function of R_k. *Hint:* See page 65 of lecture notes.

(b) Recall from page 65 that R_k/k^2 weakly converges to the Lévy distribution (defined on page 61) as $k \to \infty$. Use this to show that the characteristic function of the Lévy distribution is $\exp(-\sqrt{-2it})$. *Hint:*

$$\lim_{n\to\infty} (1 - a_n)^n = e^{-z} \quad \text{if} \quad \lim_{n\to\infty} n \cdot a_n = z$$

Solution:

(a) The generating function of R_k is $(G(z))^k$, where $G(z)$ is the generating function of R_1, since R_k is the sum of k i.i.d. copies of R_1. Now R_1 has the same distribution as $T_1 + 1$, thus

$$G(z) = \mathbb{E}(z^{R_1}) = \mathbb{E}(z^{T_1+1}) = \mathbb{E}(z^{T_1} z) = z\mathbb{E}(z^{T_1}) = z\frac{1-\sqrt{1-z^2}}{z} = 1 - \sqrt{1-z^2}$$

Thus $\mathbb{E}(z^{R_k}) = \left(1 - \sqrt{1-z^2}\right)^k$.

(b) $\mathbb{E}(e^{itR_k}) = \left(1 - \sqrt{1 - e^{2it}}\right)^k$, thus if we define $\varphi_k(t) = \mathbb{E}(e^{itR_k/k^2}) = \left(1 - \sqrt{1 - e^{2it/k^2}}\right)^k$, then we have to calculate $\varphi(t) = \lim_{k\to\infty} \varphi_k(t)$, and by the second theorem on page 91 of the scanned lecture notes, $\varphi(t)$ will be the characteristic function of the Lévy distribution. By the hint, we only need to show that

$$\lim_{k\to\infty} k \cdot \sqrt{1 - e^{2it/k^2}} = \sqrt{-\lim_{k\to\infty} \frac{e^{2it/k^2} - 1}{1/k^2}} = \sqrt{-\frac{\mathrm{d}}{\mathrm{d}x} e^{2itx}\Big|_{x=0}} = \sqrt{-2it}.$$

Thus $\varphi(t) = \exp(-\sqrt{-2it})$.

Remark: It would have been quite painful to calculate $\int_0^\infty e^{itx} f(x)\,\mathrm{d}x = \exp(-\sqrt{-2it})$ directly, where $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2x}\right) x^{-3/2}$ is the p.d.f. of the Lévy distribution.

3. (15 marks) *Coupon collector's problem.* Suppose that there is an urn with n different coupons in it. You start to draw coupons from the urn with replacement. In each round you pick each coupon with equal probability. Denote by V_n the number of coupons that you need to draw until you can say that you have touched all of the coupons at least once. The goal of this exercise is to prove the limit theorem

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{V_n - n\ln(n)}{n} \le x\right) = \exp\left(-e^{-x}\right), \quad x \in \mathbb{R}. \tag{1}$$

(a) Let $\xi_{n,1}, \xi_{n,2}, \dots, \xi_{n,n}$ denote independent random variables with distribution

$$\mathbb{P}(\xi_{n,k} = m) = \frac{k}{n}\left(\frac{n-k}{n}\right)^{m-1}, \quad m = 1, 2, \dots$$

Show that V_n has the same distribution as $\xi_{n,1} + \xi_{n,2} + \dots + \xi_{n,n}$ by giving a probabilistic meaning to the random variables $\xi_{n,1}, \xi_{n,2}, \dots, \xi_{n,n}$ in the context of coupon collection.
Hint: This is very similar to the lemma proved on page 96-97 of the scanned lecture notes.

(b) Find $\lim_{n\to\infty} \mathbb{E}\left(\frac{V_n - n\ln(n)}{n}\right)$ and $\lim_{n\to\infty} \mathrm{Var}\left(\frac{V_n - n\ln(n)}{n}\right)$.

(c) Show that for any fixed $k \in \mathbb{N}$, we have $\xi_{n,k}/n \Rightarrow \mathrm{EXP}(k)$ as $n \to \infty$ using the method of characteristic functions.

(d) Show that $\frac{V_n - \mathbb{E}(V_n)}{n} \Rightarrow Z$, where $Z + \gamma$ has standard Gumbel distribution and γ is the Euler constant.
Hint: Use the method of characteristic functions and the results proved in class (see page 95-98).

(e) Conclude the proof of the result stated in equation (1).

Solution:

(a) The distribution of $\xi_{n,k}$ is optimistic $\mathrm{GEO}(\frac{k}{n})$. When you have already collected ℓ types of coupons (where $\ell = 0, 1, \dots, n-1$) then the number of draws that you have to perform until you touch a new type of coupon has $\mathrm{OPTGEO}(\frac{n-\ell}{n})$ distribution. So let $T_{n,\ell}$ denote the number of coupons drawn until you have already collected ℓ different types of coupons. Thus

$$T_{n,0} = 0, \qquad T_{n,n} = V_n, \qquad T_{n,\ell+1} - T_{n,\ell} =: \xi_{n,n-\ell} \sim \mathrm{OPTGEO}(\frac{n-\ell}{n})$$

and clearly $T_{n,1} - T_{n,0}, T_{n,2} - T_{n,1}, \dots, T_{n,n} - T_{n,n-1}$ are independent.

(b) If $X \sim \mathrm{OPTGEO}(p)$ then $\mathbb{E}(X) = \frac{1}{p}$ and $\mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{1}{p^2} - \frac{1}{p}$.
$\xi_{n,k} \sim \mathrm{OPTGEO}(\frac{k}{n})$, thus $\mathbb{E}(\xi_{n,k}) = \frac{n}{k}$, $\mathrm{Var}(\xi_{n,k}) = \frac{n^2}{k^2} - \frac{n}{k}$.
$\mathbb{E}(V_n) = n\sum_{k=1}^n \frac{1}{k}$, $\mathrm{Var}(V_n) = n^2 \sum_{k=1}^n \frac{1}{k^2} - n\sum_{k=1}^n \frac{1}{k}$.
$\lim_{n\to\infty} \mathbb{E}\left(\frac{V_n - n\ln(n)}{n}\right) = \lim_{n\to\infty} \frac{\mathbb{E}(V_n) - n\ln(n)}{n} = \lim_{n\to\infty} \sum_{k=1}^n \frac{1}{k} - \log(n) = \gamma$ (Euler's constant).
$\lim_{n\to\infty} \mathrm{Var}\left(\frac{V_n - n\ln(n)}{n}\right) = \lim_{n\to\infty} \frac{1}{n^2}\mathrm{Var}\left(V_n\right) = \sum_{k=1}^\infty \frac{1}{k^2} - \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \frac{1}{k} = \sum_{k=1}^\infty \frac{1}{k^2} = \frac{\pi^2}{6}$.

(c) $\mathbb{E}(e^{it\xi_{n,k}}) = \sum_{m=1}^\infty e^{itm}\frac{k}{n}\left(\frac{n-k}{n}\right)^{m-1} = e^{it}\frac{k}{n}\sum_{m=1}^\infty \left(e^{it}\frac{n-k}{n}\right)^{m-1} = \frac{\frac{k}{n}e^{it}}{1 - \frac{n-k}{n}e^{it}} = \frac{ke^{it}}{n-(n-k)e^{it}}$.
$\lim_{n\to\infty} \mathbb{E}(e^{it\xi_{n,k}/n}) = \lim_{n\to\infty} \frac{ke^{it/n}}{n-(n-k)e^{it/n}} = \lim_{n\to\infty} \frac{k}{k+n(1-e^{it/n})} = \frac{k}{k-it} = (1 - \frac{it}{k})^{-1}$. This is the characteristic function of the $\mathrm{EXP}(k)$ distribution.

(d) We have already seen that $Z = \sum_{k=1}^\infty (Y_k - \mathbb{E}(Y_k)) = \sum_{k=1}^\infty (Y_k - 1/k)$, where $Y_k \sim \mathrm{EXP}(k)$.
Thus $\mathbb{E}(e^{itZ}) = \prod_{k=1}^\infty (1 - \frac{it}{k})^{-1} e^{-it/k}$. Let us define $\eta_{n,k} = (\xi_{n,k} - \mathbb{E}(\xi_{n,k}))/n$. We have

$$\lim_{n\to\infty} \mathbb{E}\left[\exp\left(it\frac{V_n - \mathbb{E}(V_n)}{n}\right)\right] = \lim_{n\to\infty}\prod_{k=1}^n \mathbb{E}(e^{it\eta_{n,k}}) \overset{(*)}{=} \prod_{k=1}^\infty \lim_{n\to\infty} \mathbb{E}(e^{it\eta_{n,k}}) \overset{(c)}{=} \mathbb{E}(e^{itZ}), \quad t \in \mathbb{R}$$

but in order to verify that the equality $(*)$ indeed holds, one needs to take the logarithm of both sides and use the dominated convergence theorem. Let's see the details. By HW6.3(c) we know that

$$|\mathbb{E}(e^{it\eta_{n,k}}) - 1| = |\mathbb{E}(e^{it\eta_{n,k}}) - 1 - it\mathbb{E}(\eta_{n,k})| \le \frac{3}{2}t^2\mathbb{E}(\eta_{n,k}^2) \le \frac{3}{2}t^2\frac{1}{k^2}$$

Note that if $|z-1| < \frac{1}{2}$ then there is a constant C such that $|\ln(z)| \le C|z-1|$. Thus $|\ln(\mathbb{E}(e^{it\eta_{n,k}}))| \le C't^2\frac{1}{k^2}$ and we obtain $\lim_{n\to\infty}\sum_{k=1}^n \ln(\mathbb{E}[e^{it\eta_{n,k}}]) = \sum_{k=1}^\infty \lim_{n\to\infty}\ln(\mathbb{E}[e^{it\eta_{n,k}}])$ by dom.conv.thm.

(e) From (b) and (d) we obtain $\lim_{n\to\infty}\mathbb{E}\left[\exp\left(it\frac{V_n - n\ln(n)}{n}\right)\right] = \mathbb{E}[e^{it(Z+\gamma)}]$, but $Z + \gamma$ has standard Gumbel distribution (see page 95-98 of the scanned lecture notes), hence (1) holds.