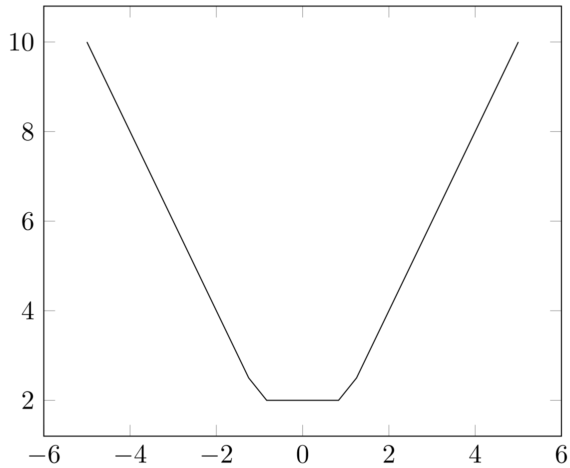
Limit / large dev. thms. first midterm practice, solutions

1. Let $f(x) = |x-1| + |x+1|$. Find the Legendre transform of f.

   Solution (by Anonymous, approved by Balázs):



$f(x)$ is the maximum of the following three linear functions:

(a) $f_1(x) = -2x$

(b) $f_2(x) = 2$

(c) $f_3(x) = 2x$

$f(x) = \max_{x\in\mathbf{R}}(f_1(x), f_2(x), f_3(x))$, this is a special case of Exercise 3. So according to this, the Legendre transform of $f(x) = \max_{x\in\mathbb{R}}(f_1(x), f_2(x), f_3(x))$ is $\hat{f}(\lambda) = \{\hat{f}_1(\lambda), \hat{f}_2(\lambda), \hat{f}_3(\lambda))\}^{co}$.

$$\hat{f}_1(\lambda) = \sup_{x\in\mathbb{R}}\{\lambda x + 2x\} = \begin{cases} \infty & \text{if } \lambda \neq -2 \\ 0 & \text{if } \lambda = -2 \end{cases}$$

$$\hat{f}_2(\lambda) = \sup_{x\in\mathbb{R}}\{\lambda x - 2\} = \begin{cases} \infty & \text{if } \lambda \neq 0 \\ -2 & \text{if } \lambda = 0 \end{cases}$$

$$\hat{f}_3(\lambda) = \sup_{x\in\mathbb{R}}\{\lambda x - 2x\} = \begin{cases} \infty & \text{if } \lambda \neq 2 \\ 0 & \text{if } \lambda = 2 \end{cases}$$

$$\boxed{\hat{f}(\lambda) = \{\hat{f}_1(\lambda), \hat{f}_2(\lambda), \hat{f}_3(\lambda))\}^{co} = \begin{cases} \infty & \text{if } \lambda < -2 \text{ or } \lambda > +2 \\ |\,\lambda\,| - 2 & \text{if } \lambda \in [-2, 2] \end{cases}}$$

2. Let $Z(\lambda)$ denote the moment generating function of the r.v. X. Denote by $X^{(\mu)}$ the exponentially tilted random variable (tilted with parameter $\mu \in \mathbb{R}$). Let $Z_\mu(\lambda) = \mathbb{E}\big(\exp(\lambda X^{(\mu)})\big)$ denote the moment generating function of $X^{(\mu)}$. Write down an identity between $Z(\lambda+\mu)$, $Z_\mu(\lambda)$ and $Z(\mu)$.

   Solution (by anonymous, approved by Balázs Ráth):

$$\mathbb{E}(g(X^{(\mu)})) = \mathbb{E}\left(g(X)\frac{e^{\mu X}}{Z(\mu)}\right) \tag{1}$$

$$Z_\mu(\lambda) = \mathbb{E}\left(\frac{e^{\lambda X}e^{\mu X}}{Z(\mu)}\right) = \frac{Z(\lambda+\mu)}{Z(\mu)} \tag{2}$$

3. If $f:\mathbb{R}\to\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}$, denote by $\{f,g\}^{co}$ the joint lower convex envelope of f and g, i.e., the supremum of those affine linear functions that lie below both f and g. Show that if f and g are both convex and continuous then the Legendre transform of $\max\{f,g\}$ is $\{\hat{f},\hat{g}\}^{co}$ and that the Legendre transform of $\{f,g\}^{co}$ is $\max\{\hat{f},\hat{g}\}$.

Solution (by Balázs):

Let us first state a lemma:

Lemma: For any function h we have $\widehat{h}=\widehat{h^{co}}$ (where h^{co} denotes the lower convex envelope of h).

Proof of lemma: observe that both functions $\widehat{h}$ and $\widehat{h^{co}}$ are convex continuous functions, so it is enough to prove that their Legendre transforms are the same. Thus it is enough to prove $\widehat{\widehat{h}}=\widehat{\widehat{h^{co}}}$, but this equality indeed holds, since both sides are equal to h^{co}. End of proof of lemma.

Now we can start solving the exercise.

We begin by proving that

$$\text{the Legendre transform of } \{f,g\}^{co} \text{ is } \max\{\hat{f},\hat{g}\}. \tag{3}$$

First note that $\{f,g\}^{co}=(f\wedge g)^{co}$. Thus we want: $\widehat{(f\wedge g)^{co}}=\hat{f}\vee\hat{g}$.

By the lemma, we want to prove: $\widehat{f\wedge g}=\hat{f}\vee\hat{g}$

We will prove this by proving $\widehat{f\wedge g}\geq\hat{f}\vee\hat{g}$ and $\widehat{f\wedge g}\leq\hat{f}\vee\hat{g}$ separately.

First we prove $\widehat{f\wedge g}\geq\hat{f}\vee\hat{g}$. Note that $f\wedge g\leq f$, thus $\widehat{f\wedge g}\geq\hat{f}$ (since the Legendre transform is order reversing). Similarly we have $\widehat{f\wedge g}\geq\hat{g}$. Thus we obtain $\widehat{f\wedge g}\geq\hat{f}\vee\hat{g}$.

Next we prove $\widehat{f\wedge g}\leq\hat{f}\vee\hat{g}$. Note that $\hat{f}\vee\hat{g}\geq\hat{f}$, thus $\widehat{\hat{f}\vee\hat{g}}\leq\hat{\hat{f}}=f$. Similarly we have $\widehat{\hat{f}\vee\hat{g}}\leq g$, thus $\widehat{\hat{f}\vee\hat{g}}\leq f\wedge g$. Taking the Legendre transform of both sides, we obtain $\widehat{\widehat{\hat{f}\vee\hat{g}}}\geq\widehat{f\wedge g}$, but $\hat{f}\vee\hat{g}$ is a convex function, thus $\widehat{\widehat{\hat{f}\vee\hat{g}}}=\hat{f}\vee\hat{g}$, thus we obtain the desired $\hat{f}\vee\hat{g}\geq\widehat{f\wedge g}$.

We are done with the proof of $\widehat{f\wedge g}=\hat{f}\vee\hat{g}$.

We are done with the proof of the first task of the exercise, i.e., that $\widehat{(f\wedge g)^{co}}=\hat{f}\vee\hat{g}$, i.e., we are done with the proof of (3).

The second task of the exercise is to prove that

$$\text{the Legendre transform of } \max\{f,g\} \text{ is } \{\hat{f},\hat{g}\}^{co}. \tag{4}$$

We will prove the second task using the first task (that we have already proved). Want: $\widehat{f\vee g}=\{\hat{f},\hat{g}\}^{co}$

Since both of the functions $\widehat{f\vee g}$ and $\{\hat{f},\hat{g}\}^{co}$ are convex and continuous, it is enough to prove that their Legendre transforms agree to show that they agree.

Thus it is enough to prove $\widehat{\widehat{f\vee g}}=\widehat{\{\hat{f},\hat{g}\}^{co}}$. Since $\widehat{\widehat{f\vee g}}=f\vee g$, its is enough to prove $f\vee g=\widehat{\{\hat{f},\hat{g}\}^{co}}$. But this identity holds by the result of the first task of the exercise (we only have to apply that formula with $\hat{f}$ in place of f and $\hat{g}$ in place of g and use that $\hat{\hat{f}}=f$ and $\hat{\hat{g}}=g$). The proof of the second task of the exercise (i.e., (4)) is complete.

4. We use a randomized algorithm to solve a yes/no decision problem. The algorithm gives the correct answer with probability $p>\frac{1}{2}$. We run the algorithm n times (where n is an odd number) and make our decision based on the majority of the results. Use the exponential Chebyshev's inequality (á la Cramér) to give a very good upper bound the probability that we make a wrong decision. Simplify the formula that you obtain as much as possible.

Solution by anonymous, approved by Balázs): Let Y_i be 1 if the ith guess is right and 0 if it is wrong. $\mathbb{P}(Y_i=1)=p>\frac{1}{2}$, $\mathbb{P}(Y_i=0)=1-p<\frac{1}{2}$. $Y_1,\dots,Y_n$ are iid, with Bernoulli(p) distribution.

$$\begin{aligned}\mathbb{P}\left(\text{wrong decision}\right)=\mathbb{P}\left(Y_1+\cdots+Y_n<\frac{n}{2}\right)=\mathbb{P}\left(\frac{Y_1+\cdots+Y_n}{n}<1/2\right)&\overset{(*)}{\leq} e^{-nI(\frac{1}{2})}\\ &=e^{-n(\frac{1}{2}\ln\left(\frac{1}{2p}\right)+\frac{1}{2}\ln\left(\frac{1}{2(1-p)}\right))}=(4p(1-p))^{\frac{n}{2}},\end{aligned}\tag{5}$$

where $(*)$ holds by the upper bound of Cramér's theorem and $I(x) = x\ln(\frac{x}{p}) + (1-x)\ln(\frac{1-x}{1-p})$ is the large deviation rate function of the Bernoulli(p) distribution.

5. Let X and Y denote independent random variables. Denote by I_X and I_Y the large deviation rate function of X and Y, respectively. Show that the large deviation function I_{X+Y} of $X+Y$ is the „infimum convolution" of I_X and I_Y.

   Hint: A non-rigorous proof using the heuristic meaning of Cramér's theorem (and our $\approxeq$ notation) is OK. A rigorous proof is even better. You should figure out by yourselves the notion of „infimum convolution" (or Google it)

   Solution: ???

6. Let X denote the random variable with p.d.f.

$$f(x) = 4xe^{-2x}\mathbb{1}[x \geq 0]$$

   Let Y denote the sum of 1000 i.i.d. copies of X.

   (a) Find the Legendre transform of the logarithmic mom.gen. function of X.

   (b) Denote by g the p.d.f. of Y. Approximate $g(1000)$.

   (c) Estimate the number of zeroes in the decimal expansion of $\mathbb{P}(Y \leq 500)$.

   Solution (by Waqar Ali Soomro, approved by Balázs):
   Let Y_1 and Y_2 be independent and identically distributed (i.i.d.) EXP(2) random variables (i.e., exponential distribution with intensity parameter $\lambda = 2$). We can define $X = Y_1 + Y_2$ since X and $Y_1 + Y_2$ have the same distribution (see Homework 3.2(a)).

   (a)

$$\begin{aligned} I_X(x) &= 2I_{Y_1}(x/2) \quad \text{(by Homework 3.1(a))} \\ &= 2(2 \cdot \frac{x}{2} - \ln(2 \cdot \frac{x}{2}) - 1) \quad \text{(because } I_{Y_1}(y) = \lambda y - \ln(\lambda y) - 1, \text{ since } Y_1 \sim \text{EXP}(\lambda)) \\ &= 2x - 2\ln(x) - 2 \end{aligned}$$

   (b) Y has the same distribution as the sum of 2000 i.i.d. copies of Y_1 where $Y_1 \sim \text{EXP}(2)$. In other words, let $Y := S_{2000} = Y_1 + \ldots + Y_{2000}$, and then g is the p.d.f. of S_{2000}. Recalling that $\mathbb{E}(Y_i) = \frac{1}{2}$ and $\text{Var}(Y_i) = \frac{1}{4}$, let us define the standardized sum

$$Y^* := \frac{S_{2000} - 2000 \cdot \frac{1}{2}}{\sqrt{2000 \cdot \frac{1}{4}}}$$

   and let g^* denote the p.d.f. of Y^*. We know from class (see page 48 of scanned lecture notes) that the p.d.f. of the standardized sum of i.i.d. exponential random variables can be approximated as follows:

$$g^*(x) \approx \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{6}$$

   Now $Y = \sqrt{500}Y^* + 1000$, thus $\mathbb{P}(Y \leq x) = \mathbb{P}(Y^* \leq \frac{x-1000}{\sqrt{500}})$ and differentiating both sides w.r.t. x we obtain $g(x) = g^*(\frac{x-1000}{\sqrt{500}}) \cdot \frac{1}{\sqrt{500}}$, thus

$$g(1000) = g^*(0) \cdot \frac{1}{\sqrt{500}} \overset{(6)}{\approx} \frac{1}{\sqrt{2\pi}\sqrt{500}}$$

   (c)

$$\mathbb{P}(Y \leq 500) = \mathbb{P}(S_{2000} \leq 500) = \mathbb{P}(\frac{S_{2000}}{2000} \leq \frac{1}{4}) \approxeq e^{-2000 \cdot \inf_{y \leq 1/4} I_{Y_1}(y)} = e^{-2000 \cdot I_{Y_1}(1/4)} = \\ e^{-2000 \cdot (2\frac{1}{4} - \ln\frac{2}{4} - 1)} = e^{-386.3} = 10^{\log_{10}(e^{-386.3})} = 10^{-168}, \tag{7}$$

   where in the triple wiggly equivalence (i.e., $\approxeq$) holds by Cramer's theorem. Thus, there are approximately 168 zeroes in the decimal expansion of $P(Y \leq 500)$ before the first non-zero digit.

7. I roll a fair die 1000 times. Denote by X the sum of the numbers rolled.

   (a) Estimate the probability that X is greater than or equal to 3550.

   (b) Estimate the probability that X is exactly equal to 3550.

   (c) Give a good lower bound on the number of zeroes in the decimal expansion of the probability that X is greater than or equal to 4500.

   Solution (by Róbert, approved by Balázs): Notice that $X = S_{1000} = \sum_{k=1}^{1000} Y_k$ where $Y_1, Y_2, \ldots$ are i.i.d. random variables with UNI$\{1, 2, \ldots, 6\}$ distribution (i.e., $\mathbb{P}(Y_i = k) = \frac{1}{6}$, $k = 1, 2, \ldots, 6$). We will need the expected value and the standard deviation of these random variables:

$$\mu = \mathbb{E}\,(Y_k) = 3.5, \tag{8}$$

$$\sigma^2 = \mathrm{Var}(Y_k) = \frac{35}{12}, \tag{9}$$

$$\sigma = \sqrt{\mathrm{Var}(Y_k)} = \sqrt{\frac{35}{12}}, \tag{10}$$

$$\mu_X = 1000 \cdot \mu = 3500, \tag{11}$$

$$Var_X = 1000 \cdot \sigma^2 = \frac{8750}{3}, \tag{12}$$

$$\sigma_X = \sqrt{Var_X} = \sqrt{1000} \cdot \sigma \tag{13}$$

   Thus $\mathbb{E}(S_{1000}) = \mathbb{E}(X) = \mu_X$ and $\mathrm{Var}(S_{1000}) = \mathrm{Var}(X) = Var_X$.

   (a) Note that 3550 is quite close to μ_X, thus in order to estimate $\mathbb{P}(S_{1000} \geq 3550)$, we will use the Central Limit Theorem (CLT):

$$\mathbb{P}(X \geq 3550) = \mathbb{P}\Big(\frac{X - \mu_X}{\sigma_X} \geq \frac{3550 - \mu_X}{\sigma_X}\Big) \approx 1 - \Phi\Big(\frac{3550 - 3500}{\sqrt{1000}\sqrt{35/12}}\Big) = 1 - \Phi(0.926) = 0.177 \tag{14}$$

   (b) The solution of this sub-exercise is not fully rigorous, because we did not prove the local CLT for the sum of i.i.d. UNI$\{1, 2, \ldots, 6\}$ random variables (for the rigorous statement and proof of the local CLT, see e.g. Section 2 of the book *Random walk: A modern introduction* by Greg Lawler and Vlada Limic (Cambridge University Press, 2010.)). Thus the following solution is heuristic, but it gives the true answer.

   Let us consider the smoothed random variable

$$X^* = X + U$$

   where $U \sim \mathrm{UNI}(0, 1)$ is independent form X: we use U to continuously fill in the gaps between the integer values that X can take. Let us also introduce the notation

$$\widehat{X} = \frac{X^* - \mu_X}{\sigma_X}$$

   where μ_X and σ_X are defined above.

   Let g^* denote the p.d.f. of X^* and let $\widehat{g}$ denote the p.d.f. of $\widehat{X}$. Note that we have

$$g^*(x) = \mathbb{P}(X = \lfloor x \rfloor), \qquad g^*(x) = \widehat{g}\left(\frac{x - \mu_X}{\sigma_X}\right) \cdot \frac{1}{\sigma_X} \tag{15}$$

   Also, by the CLT and Slutsky, we can approximate $\widehat{X}$ by $\mathcal{N}(0, 1)$, and therefore (heuristically!) we can approximate the p.d.f. of $\widehat{X}$ by the p.d.f. of $\mathcal{N}(0, 1)$:

$$\widehat{g}(x) \approx \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{16}$$

   This approximation works if $|x|$ is not too big. This is an example of the local CLT: we have proved similar results in the case when the summands have exponential distribution (see page 48 of the scanned lecture notes) and in the case when the summands have BER(1/2) distribution (see HW4.3

and pages 55-56 of scanned). However, we did not prove the local CLT for sums of fair die rolls, but the local CLT does hold in this case, see the above cited book of Lawler and Limic.
The answer to the question of this sub-exercise is

$$\mathbb{P}(X=3550) \overset{(15)}{=} g^*(3550) \overset{(15)}{=} \widehat{g}\left(\frac{3550-3500}{\sqrt{1000}\sqrt{35/12}}\right)\cdot\frac{1}{\sqrt{1000}\sqrt{35/12}} = \widehat{g}(0.926)\cdot\frac{1}{\sigma\sqrt{1000}} \overset{(16)}{\approx} \frac{1}{\sqrt{2\pi}}e^{-(0.926)^2/2}\frac{1}{\sigma\sqrt{1000}}. \tag{17}$$

(c) Our first instinct tells us to use Cramer's theorem, but the problem is that we cannot express the large deviation rate function $I(x)$ of the UNI$\{1,2,...,6\}$ distribution using elementary functions (the logarithmic moment generating function of UNI$\{1,2,...,6\}$ is $\widehat{I}(\lambda)=\ln\left(\frac{1}{6}\sum_{k=1}^{6}e^{\lambda k}\right)$, but the Legendre transform of this function is ugly). So, for this sub-exercise, we use Hoeffding's inequality which states that

$$\mathbb{P}(S_n \geq \mathbb{E}(S_n)+t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i-a_i)^2}\right) \tag{18}$$

holds if $S_n=\sum_{i=1}^{n}Y_i$ where Y_i are independent and $\mathbb{P}(Y_i\leq b_i)=\mathbb{P}(Y_i\geq a_i)=1$. In our case $t=1000$ and $a_i=1$, $b_i=6$ and thus $b_i-a_i=5$ for all i. Substituting these into Hoeffding's inequality we can conclude that:

$$\mathbb{P}(X\geq 4500)\leq e^{-80}\approx 1.8\cdot 10^{-35} \tag{19}$$

which means that there are at least 34 zero digits before the first non-zero digit.

8. Let $X_1, X_2, \ldots$ denote i.i.d. r.v.'s with POI(λ) distribution. Let $S_n = X_1+\cdots+X_n$.

(a) Use Stirling's formula to prove the *local CLT* for S_n:

$$\lim_{n\to\infty}\sqrt{n\lambda}\,\mathbb{P}\left(S_n=\lfloor n\lambda+\sqrt{n\lambda}x\rfloor\right)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}. \tag{20}$$

(b) Deduce the *global CLT* from the local CLT: show that

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{S_n-\mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}\leq x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,\mathrm{d}y.$$

Solution (by Balázs):

(a) We fix the value of $\lambda>0$ and $x\in\mathbb{R}$. We write $a_n\approx b_n$ to denote that $\lim_{n\to\infty}a_n/b_n=1$. We will first show that if $k(n)\in\mathbb{N}$ for each n such that $k(n)=n\lambda+\sqrt{n\lambda}z(n)$, where $z(n)=z(n,\lambda,x)$ is a bounded sequence then

$$\sqrt{n\lambda}\,\mathbb{P}\left(S_n=n\lambda+\sqrt{n\lambda}z(n)\right)\approx\frac{1}{\sqrt{2\pi}}e^{-z(n)^2/2}. \tag{21}$$

Indeed,

$$\sqrt{n\lambda}\,\mathbb{P}\left(S_n=n\lambda+\sqrt{n\lambda}z(n)\right)=\sqrt{n\lambda}\,\mathbb{P}\left(S_n=k(n)\right)\overset{(*)}{=}\sqrt{n\lambda}e^{-n\lambda}\frac{(n\lambda)^{k(n)}}{k(n)!}\overset{(**)}{\approx} \sqrt{n\lambda}e^{-n\lambda}\frac{(n\lambda)^{k(n)}}{\sqrt{2\pi}\sqrt{k(n)}k(n)^{k(n)}e^{-k(n)}}=\frac{1}{\sqrt{2\pi}}\left(1+\frac{z(n)}{\sqrt{n\lambda}}\right)^{-1/2}\left(1+\frac{z(n)}{\sqrt{n\lambda}}\right)^{-k(n)}e^{\sqrt{n\lambda}z(n)}\approx \frac{1}{\sqrt{2\pi}}\left(1+\frac{z(n)}{\sqrt{n\lambda}}\right)^{-k(n)}e^{\sqrt{n\lambda}z(n)}, \tag{22}$$

where equation $(*)$ holds since $S_n\sim$ POI$(n\lambda)$ and the approximate equation $(**)$ holds by Stirling. Multiplying by $\sqrt{2\pi}$ and taking the logarithm, we want to show

$$\ln\left(\left(1+\frac{z(n)}{\sqrt{n\lambda}}\right)^{-k(n)}e^{\sqrt{n\lambda}z(n)}\right)=-z(n)^2/2+o(1), \tag{23}$$

where $o(1)$ denotes an error term that goes to zero as $n \to \infty$. We have

$$\ln\left(\left(1+\frac{z(n)}{\sqrt{n\lambda}}\right)^{-k(n)} e^{\sqrt{n\lambda}z(n)}\right) = -k(n)\ln\left(1+\frac{z(n)}{\sqrt{n\lambda}}\right) + \sqrt{n\lambda}z(n) \stackrel{(\bullet)}{=}$$

$$-k(n)\left(\frac{z(n)}{\sqrt{n\lambda}} - \frac{1}{2}\frac{z(n)^2}{n\lambda} + O(n^{-3/2})\right) + \sqrt{n\lambda}z(n) =$$

$$-(n\lambda + \sqrt{n\lambda}z(n))\left(\frac{z(n)}{\sqrt{n\lambda}} - \frac{1}{2}\frac{z(n)^2}{n\lambda} + O(n^{-3/2})\right) + \sqrt{n\lambda}z(n) =$$

$$-(\sqrt{n\lambda} + z(n))\left(z(n) - \frac{1}{2}\frac{z(n)^2}{\sqrt{n\lambda}} + O(n^{-1})\right) + \sqrt{n\lambda}z(n) =$$

$$-\sqrt{n\lambda}z(n) + \frac{1}{2}z(n)^2 + O(n^{-1/2}) - z(n)^2 + O(n^{-1/2}) + \sqrt{n\lambda}z(n) = -\frac{1}{2}z(n)^2 + o(1), \quad (24)$$

where in $(\bullet)$ we used that $\ln(1+y) = y - y^2/2 + O(|y|^3)$. The proof of (23) is complete. The proof of (21) is complete.

Now it is easy to prove (20). Let us define

$$z(n) = \frac{\lfloor n\lambda + \sqrt{n\lambda}x\rfloor - n\lambda}{\sqrt{n\lambda}}, \quad (25)$$

so that we have $\lfloor n\lambda + \sqrt{n\lambda}x\rfloor = n\lambda + \sqrt{n\lambda}z(n)$. We have $\lim_{n\to\infty} z(n) = x$, thus in particular $(z(n))$ is a bounded sequence. We obtain the desired

$$\sqrt{n\lambda}\mathbb{P}\left(S_n = \lfloor n\lambda + \sqrt{n\lambda}x\rfloor\right) \stackrel{(21)}{\approx} \frac{1}{\sqrt{2\pi}}e^{-z(n)^2/2} \approx \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

The proof of (20) is complete.

(b) The usual Scheffé and Slutsky tricks, similarly to pages 55-56 of scanned lecture notes.

9. Before the national election, we want to estimate the fraction p of republican voters. We ask n random people and calculate the fraction p_n of republicans in our sample. How to choose n if we want to estimate the value of p with a margin of error 0.01 with 95% confidence? Use the CLT.

Solution: ???

10. *Weibull distribution.* Let $U_1, U_2, \ldots$ denote i.i.d. random variables with UNI$[0,1]$ distribution. Let $\beta > 0$. Let

$$M_n = \min\{U_1^\beta, \ldots, U_n^\beta\}.$$

Show that $n^\alpha M_n$ converges in distribution as $n \to \infty$ to a non-trivial probability distribution if we choose $\alpha > 0$ correctly. Determine the cumulative distribution function (c.d.f.) $F(x)$ of the limiting distribution.

Solution: Jocó (approved by Balázs)

$$\begin{aligned}\mathbb{P}\left(n^\alpha \min(U_1^\beta, \ldots, U_n^\beta) \le x\right) &= 1 - \mathbb{P}\left(\min(U_1^\beta, \ldots, U_n^\beta) \ge \frac{x}{n^\alpha}\right) \\ &= 1 - \mathbb{P}\left(U_1^\beta \ge \frac{x}{n^\alpha}\right)^n \\ &= 1 - \mathbb{P}\left(U_1 \ge \left(\frac{x}{n^\alpha}\right)^{1/\beta}\right)^n \\ &= 1 - \left(1 - \left(\frac{x}{n^\alpha}\right)^{1/\beta}\right)^n\end{aligned} \quad (26)$$

Thus if we choose $\alpha := \beta$, then we have that:

$$\mathbb{P}\left(n^\alpha \min(U_1^\beta, \ldots, U_n^\beta) \le x\right) \to 1 - e^{-x^{1/\beta}} \quad as \quad n \to \infty.$$

Therefore the CDF $F(x)$ of the weak limit of $n^\alpha M_n$ is:

$$F(x) = \left(1 - e^{-x^{1/\beta}}\right)\mathbb{1}[x \ge 0].$$

11. Let $\tau_1, \tau_2, \ldots$ be i.i.d. waiting times between successive events and define the *renewal process*

$$\nu_t := \max\left\{ n \, : \, \sum_{i=1}^{n} \tau_i < t \right\}.$$

In plain words, ν_t is the number of events that occurred in the time interval $[0, t]$. Denote $m := \mathbb{E}(\tau_j) < \infty$, $\sigma^2 := \mathrm{Var}(\tau_j) < \infty$. Use the classic CLT for the sum of i.i.d. r.v.'s to derive a CLT for ν_t: find $a > 0, b > 0$ such that

$$\lim_{t\to\infty} \mathbb{P}\Big(\frac{\nu_t - at}{b\sqrt{t}} < x\Big) = \Phi(x),$$

where Φ is the standard normal c.d.f. Express a and b in terms of m and σ.

Solution by Levente (approved by Balázs):

First notice that if $k \in \mathbb{R}$ then

$$\{\nu_t < k\} = \{\nu_t < \lceil k \rceil\} = \Big\{ \sum_{i=1}^{\lceil k \rceil} \tau_i > t \Big\} = \Big\{ \sum_{i=1}^{\lceil k \rceil} \tau_i \le t \Big\}^c$$

which has the exact form for the CLT for the sum random iid variables τ_i. Using this and remembering that $\mathbb{E}(\sum_{i=1}^{k} \tau_i) = mk$, $\mathrm{Var}(\sum_{i=1}^{k} \tau_i) = k\sigma^2$, we obtain that for any $a, b \in \mathbb{R}_+$ we have

$$\mathbb{P}\Big(\frac{\nu_t - at}{b\sqrt{t}} < x\Big) = \mathbb{P}\big(\nu_t < xb\sqrt{t} + at\big) = \mathbb{P}\Big(\Big\{ \sum_{i=1}^{\lceil xb\sqrt{t}+at \rceil} \tau_i \le t \Big\}^c\Big) = 1 - \mathbb{P}\Big(\sum_{i=1}^{\lceil xb\sqrt{t}+at \rceil} \tau_i < t \Big) =$$

$$= 1 - \mathbb{P}\left(\frac{\big(\sum_{i=1}^{\lceil xb\sqrt{t}+at \rceil} \tau_i\big) - m \cdot \lceil xb\sqrt{t} + at \rceil}{\sigma \cdot \sqrt{\lceil xb\sqrt{t} + at \rceil}} < \frac{t - m \cdot \lceil xb\sqrt{t} + at \rceil}{\sigma \cdot \sqrt{\lceil xb\sqrt{t} + at \rceil}} \right).$$

This at the first hand may look dangerous but observe that random variable on the left-hand side of the inequality inside the $\mathbb{P}$ can be substituted by a standard normal by the CLT. For the right side, remember HW4.2(a), which we can apply if the distribution of where we converge has a continuous c.d.f. (Φ is continuous) and if the right-hand side converges as a sequence of real numbers as t approaches ∞. This will determine a and b for us:

$$-x \longleftarrow \frac{t - m \cdot \lceil xb\sqrt{t} + at \rceil}{\sigma \cdot \sqrt{\lceil xb\sqrt{t} + at \rceil}} \iff -x \longleftarrow \frac{t - m \cdot (xb\sqrt{t} + at)}{\sigma \cdot \sqrt{xb\sqrt{t} + at}} = \frac{t(1 - ma) - mxb\sqrt{t}}{\sigma \cdot \sqrt{xb\sqrt{t} + at}}$$

The t in the numerator will overpower both $-\sqrt{t}$ and $1/\sqrt{t}$, so we need $1 - am = 0$ to the convergence, yielding us $a := m^{-1}$. Finally consider

$$\frac{-mxb\sqrt{t}}{\sigma \cdot \sqrt{xb\sqrt{t} + at}} = \frac{-mxb}{\sigma \cdot \sqrt{xb\frac{\sqrt{t}}{t} + a}} \longrightarrow \frac{-mxb}{\sigma\sqrt{a}}.$$

Choosing $b := \sigma/ma^{1/2} = \sigma m^{-3/2}$, applying HW4.2(a) we get

$$\lim_{t\to\infty} \mathbb{P}\Big(\frac{\nu_t - m^{-1}t}{\sigma m^{-3/2}\sqrt{t}} < x\Big) = 1 - \Phi(-x) = \Phi(x)$$

for all $x \in \mathbb{R}$. This result is called the CLT for renewal processes. The times $\tau_1, \tau_1 + \tau_2, \tau_1 + \tau_2 + \tau_3, \ldots$ are the renewal times (e.g. the times when we change the light bulb), ν_t is the number of light bulb changes up to time t).

12. Let $X_1, X_2, \ldots$ denote i.i.d. r.v.'s and assume that $\mathbb{E}(X_i) = 0$ and $\mathbb{P}(|X_i| \le K) = 1$ for some $K \in \mathbb{R}$. Let us define

$$Y_n = \prod_{k=1}^{n} \left(1 + \frac{X_k}{\sqrt{n}}\right).$$

Use the CLT to show that Y_n converges weakly as $n \to \infty$. The limiting distribution is famous (e.g., in financial mathematics): name it and identify its parameter(s).

Solution by Dani, streamlined by Balázs:
We shall take the logarithm of both sides.

$$\log Y_n = \sum_{k=1}^{n} \log\left(1 + \frac{X_k}{\sqrt{n}}\right) \tag{27}$$

The main idea is to use the second order Taylor expansion of $\log(1+x)$ so we end up with a sum of i.i.d. rws. We have

$$\log(1+x) = x - x^2/2 + \epsilon(x), \qquad |\epsilon(x)| \le Cx^3, \text{ if } |x| \le 1/2 \tag{28}$$

Applying this to the r.h.s. of (27), we obtain

$$\sum_{k=1}^{n} \log\left(1 + \frac{X_k}{\sqrt{n}}\right) = \sum_{k=1}^{n} \frac{X_k}{\sqrt{n}} - \frac{1}{2}\sum_{k=1}^{n} \frac{X_k^2}{n} + \sum_{k=1}^{n} \epsilon\left(\frac{X_k}{\sqrt{n}}\right) \tag{29}$$

We will deal with the three sums on the r.h.s. separately.
For the first term we can use CLT to deduce $\sum_{k=1}^n \frac{X_k}{\sqrt{n}} \Rightarrow \mathcal{N}(0,\sigma^2)$, where $\sigma^2 = \mathrm{Var}(X_i)$.

For the second term, we have $-\frac{1}{2}\sum_{k=1}^n \frac{X_k^2}{n} \Rightarrow -\frac{\sigma^2}{2}$ by the weak law of large numbers (recalling that $E(X_i) = 0$ implies $E(X_i^2) = \sigma^2$).

The third term goes to zero, since $|\epsilon\left(\frac{X_k}{\sqrt{n}}\right)| \le CK^3/n^{3/2}$ (if n is large enough so that $K/\sqrt{n} \le 1/2$) by (28), and $\sum_{k=1}^n CK^3/n^{3/2} = CK^3/\sqrt{n} \to 0$ as $n \to \infty$.

Putting together these three observations about the r.h.s. of (29), Slutsky's lemma provides

$$\log Y_n \Rightarrow \mathcal{N}\left(-\frac{\sigma^2}{2}, \sigma^2\right),$$

therefore $Y_n \Rightarrow e^X$, where $X \sim \mathcal{N}\left(-\frac{\sigma^2}{2}, \sigma^2\right)$, thus the weak limit of Y_n follows a log-normal distribution.

13. Let $X_0^n, X_1^n, X_2^n, \ldots, X_{2n}^n$ denote the conditional distribution of one dimensional simple symmetric random walk under the condition that it returns to the origin in $2n$ steps, i.e., that $X_{2n}^n = 0$. Denote by $M_n = \max\{X_0^n, X_1^n, X_2^n, \ldots, X_{2n}^n\}$. Show that $M_n/\sqrt{n}$ converges in distribution as $n \to \infty$ and find the c.d.f. of the limiting distribution.

Solution (by Balázs): We will first show that

$$\mathbb{P}(M_n \ge k) = \frac{\binom{2n}{n+k}2^{-2n}}{\binom{2n}{n}2^{-2n}}, \qquad k = 0, 1, \ldots, n \tag{30}$$

Let (X_n) denote a simple random walk. Let T_k denote the hitting time of level k. Note that we have

$$\mathbb{P}(M_n \ge k) = \mathbb{P}(T_k \le 2n \,|\, X_{2n} = 0) = \frac{\mathbb{P}(T_k \le 2n,\, X_{2n} = 0)}{\mathbb{P}(X_{2n} = 0)} \stackrel{(*)}{=} \frac{\mathbb{P}(T_k \le 2n,\, X_{2n} = 2k)}{\binom{2n}{n}2^{-2n}} = \frac{\mathbb{P}(X_{2n} = 2k)}{\binom{2n}{n}2^{-2n}} \stackrel{(**)}{=} \frac{\binom{2n}{n+k}2^{-2n}}{\binom{2n}{n}2^{-2n}}, \tag{31}$$

where $(*)$ holds by the reflection principle and $(**)$ holds because $\{X_{2n} = 2k\}$ holds if and only if exactly $n+k$ increments (out of $2n$) are „up" and exactly $n-k$ increments (out of $2n$) are „down".

We are done with the proof of (30). Next we show that

$$\lim_{n\to\infty} \mathbb{P}(M_n/\sqrt{n} \ge y) = e^{-y^2}, \qquad y \ge 0. \tag{32}$$

Recall from HW4.3 that if $S_n \sim \mathrm{BIN}(n, 1/2)$ then

$$\lim_{n\to\infty} \frac{\sqrt{n}}{2}\mathbb{P}\left(S_n = \left\lfloor \frac{n}{2} + \frac{\sqrt{n}}{2}x \right\rfloor\right) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}. \tag{33}$$

Replacing n by $2n$ and $x = \sqrt{2}y$ in the above formula, we have

$$\lim_{n\to\infty} \frac{\sqrt{n}}{\sqrt{2}} \mathbb{P}\left(S_{2n} = n + \lfloor \sqrt{n} y \rfloor\right) = \frac{1}{\sqrt{2\pi}} e^{-y^2}. \tag{34}$$

Noting that $\mathbb{P}(S_{2n} = n+k) = \binom{2n}{n+k} 2^{-2n}$, we obtain

$$\lim_{n\to\infty} \mathbb{P}(M_n/\sqrt{n} \ge y) = \lim_{n\to\infty} \mathbb{P}(M_n \ge \lfloor \sqrt{n} \cdot y \rfloor) \overset{(30)}{=} \lim_{n\to\infty} \frac{\mathbb{P}\left(S_{2n} = n + \lfloor \sqrt{n} y \rfloor\right)}{\mathbb{P}(S_{2n} = n+0)} \overset{(34)}{=} \frac{\frac{1}{\sqrt{2\pi}} e^{-y^2}}{\frac{1}{\sqrt{2\pi}} e^{-0^2}} = e^{-y^2}. \tag{35}$$

We are done with the proof of (32). Thus $M_n/\sqrt{n}$ weakly converges as $n \to \infty$ to a random variable with c.d.f. $F(y) = 1 - e^{-y^2}$ for $y \ge 0$ (and $F(y) = 0$ for $y \le 0$).

14. For $n, m \in \mathbb{N}$ let X_n, Y_m be independent r.v.'s with distributions $X_n \sim \text{POI}(n)$, $Y_m \sim \text{POI}(m)$. Prove that

$$\frac{X_n - n - (Y_m - m)}{\sqrt{X_n + Y_m}}$$

converges in distribution as $n, m \to \infty$. Identify the limiting distribution.

Hint: This is easy if you use Slutsky in a clever way, similarly to page 54 of the scanned lecture notes.

Solution: ???

15. Let $X_1, X_2, \dots$ denote i.i.d. random variables with p.d.f. $f(x) = \frac{1}{(x-1)^2} \mathbb{1}[x \le 0]$. Let $M_n = \max\{X_1, \dots, X_n\}$. Find a sequence (a_n) such that M_n/a_n weakly converges to a non-degenerate random variable Z. Find the c.d.f. of Z. What does it mean that Z is max-stable?
Solution (by Péter Elek, approved by Balázs): Let $x < 0$ and $a_n > 0$. For every $n \in \mathbb{N}$ we have

$$\mathbb{P}\left(\frac{M_n}{a_n} < x\right) = \mathbb{P}(M_n < a_n x) = \mathbb{P}(X_1 < a_n x) \cdot \dots \cdot \mathbb{P}(X_n < a_n x) = [\mathbb{P}(X_1 < a_n x)]^n$$

$$\mathbb{P}(X_1 < x) = \int_{-\infty}^{x} \frac{1}{(x-1)^2} dx = \frac{1}{1-x}$$

From these we obtain

$$\mathbb{P}\left(\frac{M_n}{a_n} < x\right) = \left[\frac{1}{1 - a_n x}\right]^n = [1 - a_n x]^{-n}$$

If $a_n = 1/n$ then $\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n}{a_n} < x\right) = e^x$ for any $x \le 0$. Thus $n \cdot M_n \Rightarrow Z$, where the c.d.f. of Z is $F(x) = \min\{1, e^x\}$ for all $x \in \mathbb{R}$. Thus $-Z \sim \text{EXP}(1)$, in words: minus Z has exponential distribution with parameter 1.

Z is max-stable because if we take the maximum of i.i.d. random variables with the same distribution as Z then the resulting random variable has the same distribution as Z, up to a scaling by a constant. With formula:

$$Z_1, \dots Z_n \quad \text{i.i.d. and} \quad Z_i \sim Z \quad \text{then} \quad M_n = \max\{Z_1, \dots, Z_n\} \sim Z/n$$

This was an exercise in extreme value theory, and the limiting distribution (i.e., the distribution of Z) is a special case of the Weibull distribution, see exercise 0.10 above.

16. At time zero a stock broker has 1000 dollars. At each time-step, three things can happen: either he gains one dollar (this happens with probability 1/4), loses one dollar (this also happens with probability 1/4) or he neither gains nor loses (this happens with probability 1/2).

How would you approximate the distribution of the time when the stock broker goes bankrupt (i.e., loses all his money)? How to make this rigorous?

Solution (sketch proof only): Let $\hat{X}_n$ denote the amount of dollars our stock broker has at time n. We have $\hat{X}_n = 1000 + \frac{1}{2} X_{2n}$, where (X_n) is a simple symmetric random walk that starts at the origin. Thus the time when $(\hat{X}_n)$ hits zero can be related to the first time when the random walk (X_n) hits level -2000. We have proved limit theorems about such hitting times in class (the Lévy distribution is the limiting distribution of scaled random walk hitting times)

17. Let Z_n denote an integer-valued random variable for which

$$\mathbb{P}(Z_n = k) = (k+1)\frac{1}{n^2}(1-\frac{1}{n})^k, \qquad k = 0, 1, 2, \ldots$$

Show that Z_n/n converges in distribution as $n \to \infty$ and identify the limiting distribution.

Solution (by Balázs): Let $Y_n = Z_n + U$, where $U \sim \text{UNI}[0,1]$. The p.d.f. of Y_n is $g_n(x) = \mathbb{P}(Z_n = \lfloor x \rfloor)$. The p.d.f. of Y_n/n is

$$f_n(x) := n \cdot g_n(n \cdot x) = n \cdot \mathbb{P}(Z_n = \lfloor n \cdot x \rfloor) = \frac{\lfloor n \cdot x \rfloor + 1}{n}(1 - \frac{1}{n})^{\lfloor n \cdot x \rfloor} \to x e^{-x}, \quad n \to \infty$$

for $x \geq 0$ and of course $f_n(x) = 0$ if $x < 0$. Thus by Scheffé we have $Y_n/n \Rightarrow Y$, where the p.d.f. of Y is $f(x) = xe^{-x}\mathbb{1}[x \geq 0]$. We have $Z_n/n = Y_n/n - U/n$ and $U/n \Rightarrow 0$, so by Slutsky we have $Z_n/n \Rightarrow Y$.

18. Let X_1 and X_2 be i.i.d. random variables with Lévy distribution. What is the distribution of $X_1 + 3X_2$?

Solution (by Lili), approved by Balázs: We know that if X_1 and X_2 are i.i.d. random variables with Lévy distribution, then $X_1 + X_2$ has the same distribution as $4X_1$ (cf. page 64 of scanned lecture notes). Therefore, we suspect that $X_1 + 3X_2$ will also have Lévy distribution.

Using the same argument as before (see page 62-63 of scanned), for any $a \in \mathbb{R}_+$, let us define

$$S_{\lfloor an \rfloor} := \eta_1 + \cdots + \eta_{\lfloor an \rfloor},$$

where $\eta_1, \eta_2, \ldots$ are i.i.d. random variables copies of T_1. We know that: $\frac{S_{\lfloor an \rfloor}}{\lfloor an \rfloor^2} \Rightarrow X$, where X has Lévy distribution. Rearranging this a bit, we obtain

$$\frac{S_{\lfloor an \rfloor}}{n^2} \Rightarrow a^2 X. \tag{36}$$

We split the sums of the η's into two independent sums:

$$S_{\lfloor \sqrt{3}n \rfloor} + S_n^* = S_{\lfloor \sqrt{3}+1 \rfloor n}, \tag{37}$$

where $S_{\lfloor \sqrt{3}n \rfloor} = \eta_1 + \cdots + \eta_{\lfloor \sqrt{3}n \rfloor}$ and $S_n^* = \eta_{\lfloor \sqrt{3}n \rfloor + 1} + \cdots + \eta_{\lfloor (\sqrt{3}+1)n \rfloor}$. Note that $S_{\lfloor \sqrt{3}n \rfloor}$ and S_n^* are independent. Also note that $S_n^* \sim S_n$. Applying (36) three times, we obtain

$$\frac{S_{\lfloor \sqrt{3}n \rfloor}}{n^2} \Rightarrow 3X_1, \qquad \frac{S_n^*}{n^2} \Rightarrow X_2, \qquad \frac{S_{\lfloor (\sqrt{3}+1)n \rfloor}}{n^2} \Rightarrow (\sqrt{3}+1)^2 X, \tag{38}$$

where X_1, X_2 and X have Lévy distribution, moreover X_1 and X_2 are independent. Thus, by (37) we obtain

$$3X_1 + X_2 \sim (\sqrt{3}+1)^2 X$$

More generally, for any $a, b \in \mathbb{R}_+$, if X_1, X_2 are i.i.d. with Lévy distribution then

$$a \cdot X_1 + b \cdot X_2 \sim (\sqrt{a} + \sqrt{b})^2 \cdot X,$$

where X has Lévy distribution.