Limit / large dev. thms. first midterm practice

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

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 $Z(\mu)$.

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 $\max\{\hat{f},\hat{g}\}$.

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.

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)

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 $\mathbb{P}(Y \leq 500)$.

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.

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:

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

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

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

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.

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

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

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.

11. Let $\tau_1, \tau_2, \dots$ 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.

12. Let $X_1, X_2, \dots$ 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).

13. Let $X_0^n, X_1^n, X_2^n, \dots, X_{2n}^n$ 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 $M_n = \max\{X_0^n, X_1^n, X_2^n, \dots, X_{2n}^n\}$. Show that $M_n/\sqrt{n}$ converges in distribution as $n \to \infty$ and find the c.d.f. of the limiting distribution.

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

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?

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?

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

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

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

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