Final Exam

Problem 1.1 (Doob's lemma)**.** Let X be a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} = \sigma(X)$ be the σ-algebra generated by X. Show that for any random variable Y, measurable with respect to $\mathcal{G}$, there exists a Borel function $h : \mathbb{R} \to \mathbb{R}$ such that $Y = h(X)$.

Problem 1.2 (Barndorff-Nielsen's extension of the Borel-Cantelli lemma)**.** Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of events.

1. Show that $(\limsup A_n) \cap (\limsup A_n^c) \subseteq \limsup(A_n \cap A_{n+1}^c)$.
2. If $\liminf_{n\to\infty} \mathbb{P}[A_n] = 0$ and $\sum_n \mathbb{P}[A_n \cap A_{n+1}^c] < \infty$, show that $\mathbb{P}[\limsup_n A_n] = 0$.

Problem 1.3 (A criterion for membership in $\mathbb{L}\log\mathbb{L}$)**.** Let X be a nonnegative random variable, and let F be its cumulative distribution function (cdf). Show that

$$\mathbb{E}\big[X \log^+(X)\big] < \infty \text{ if and only if } \int_1^\infty \int_1^\infty \Big(1 - F(uv)\Big)\, du\, dv < \infty,$$

where $\log^+(x) = \max(\log(x), 0)$.

Problem 1.4 (The "Chi-squared" and "Student's t" distributions)**.** Let $\{X_k\}_{k\in\mathbb{N}}$ be an iid sequence of standard normal random variables.

1. Given $d \in \mathbb{N}$, the distribution of the random variable $X_1^2 + \cdots + X_d^2$ is called the **chi-squared distribution with d degrees of freedom**, denoted by $\chi^2(d)$. Compute its pdf.

   Hint: Use the convolutional identity $g_\alpha * g_\beta = g_{\alpha+\beta}$, where $g_\alpha(x) = \frac{x^{\alpha-1}}{2^\alpha \Gamma(\alpha)} \exp(-x/2) 1_{\{x>0\}}$ and Γ is the Gamma function.

2. For $n \in \mathbb{N}$, let X be the random (row) vector $X = (X_1, \dots, X_n)$ and let M be a $n \times n$ symmetric matrix such that $M^2 = M$. What is the distribution of XMX^T.

   Hint: Use the properties of the multivariate normal from Problem 5.1 in HW5.

3. For $n \geq 2$, what is the joint distribution of $Q^2 := \sum_{i=1}^n (X_i - \bar{X})^2$ and $\bar{X} := \frac{1}{n}(X_1 + \cdots + X_n)$?

   Hint: Same hint as in 2. above.

4. Show that there exists a constant C', which depends only on n, such that the pdf of the random variable $T = \frac{\sqrt{n}\bar{X}}{\sqrt{Q^2/(n-1)}}$ is given by

$$f_T(t) = C'\big(1 + t^2/d\big)^{-(d+1)/2} \text{ where } d = n - 1.$$

   Note: The distribution of T is called the **Student's t distribution with d degrees of freedom** and is denoted by $t(d)$. The value of the constant C' turns out to be $\frac{\Gamma((d+1)/2)}{\sqrt{2\pi d}\Gamma(d/2)}$. The only reason we use both $d = n - 1$ and n is to be consistent with the standard terminology.

Note: Look up the "Student's t-test" if you are curious about the significance of this problem in statistics.

Problem 1.5 (A probabilistic proof of Stirling's formula)**.** Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence with the Poisson(λ) distribution, i.e., $\mathbb{P}[X_1 = k] = e^{-\lambda}\lambda^k/k!$ for $k \in \mathbb{N}_0$.

1. What is the distribution of $Y_n = \sum_{k=1}^n X_k$, for $n \in \mathbb{N}$?
2. Set $\lambda = 1$ and let $Z_n = \frac{1}{\sqrt{n}} Y_n - \sqrt{n}$. Without evaluating it, show that $\mathbb{E}[|Z_n|]$ admits a limit and identify it. *Hint:* Use the fact that, in this case, the function $x \mapsto |x|$ can be used to "test" weak convergence, as if it belonged to $C_b(\mathbb{R})$. Prove this for extra credit.
3. Evaluate $\mathbb{E}[|Z_n|]$ explicitly and derive Stirling's formula $\lim_{n\to\infty} \frac{n!}{(n/e)^n\sqrt{2\pi n}} = 1$.

Problem 1.6 (Two exercises in conditional expectation)**.**

1. Given an example of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X \in \mathbb{L}^1$ and two sub-σ-algebras $\mathcal{G}$ and $\mathcal{H}$ of $\mathcal{F}$, such that

$$\mathbb{P}\Big[\mathbb{E}\big[\mathbb{E}[X \mid \mathcal{G}] \bigm| \mathcal{H}\big] = \mathbb{E}[X \mid \mathcal{H}]\Big] < 1.$$

2. For $X, Y \in \mathcal{L}^2$ and a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, show that the following "self-adjointness" property holds

$$\mathbb{E}\Big[X\,\mathbb{E}[Y \mid \mathcal{G}]\Big] = \mathbb{E}\Big[\mathbb{E}[X \mid \mathcal{G}]\,Y\Big].$$