PROBABILITY THEORY

MATH 154

Homework 5

Tail algebra

Problem 5.1: Bond percolation in 3 dimensions. Ω is the set of all subgraphs of the lattice $\mathbb{Z}^3$ with nearest neighbor connections. Look at σ-algebras $\mathcal{A}_e$ generated by the random variable $X_e(\omega) = 1_{\{e \in E(\omega)\}}$. The assumption $\mathrm{P}[\{X_e = 1\}] = p, \mathrm{P}[\{X_e = 0\}]$ defines a probability space in which $\{X_e\}_{e \in E}$ are independent.
a) What theorem does assure that we have a probability measure on Ω that is translation invariant?
b) The event A consists of all graphs for which there is an infinite cluster. Verify that $\mathrm{P}_p[A] \leq \mathrm{P}_q[A]$ if $p \leq q$.
c) Conclude there is a threshold p_c so that $p > p_c$ gives an infinite cluster and $p < p_e$ none with probability 1.
d) Hit the literature: what is currently the best estimate for p_c?

Jensen

Problem 5.2: a) Formulate Jensen inequality in the case $f(x) = |x|$ and show that it implies the calculus identity $|\int_0^1 f(x)\, dx| \leq \int_0^1 |f(x)|\, dx$ for a continuous function on $[0, 1]$.
b) It implies the geometric-arithmetic mean inequality $\sqrt{ab} \leq (a + b)/2$.
c) Jensen's inequality can explain risk aversion and motivate portfolio optimization. Let ϕ be a concave utility function. ($-\phi$ is convex). What does Jensen tell you about the expected utility?

Entropy

Problem 5.3: We study entropy $S(\mathcal{A})$ calculus for a finite σ-algebra.
a) Single variable: $f(x) = x \log(1/x)$ is concave. The limit $f(0) = 0$ exists.
b) Multi: the uniform distribution on $\{1, \dots, n\}$ has maximal entropy.
c) Let $\mathcal{A}_X$ be the σ-algebra of a random variable $X \in \mathcal{S}$ and $\mathcal{A}_{X,Y}$ the σ algebra of two random variables $X, Y \in \mathcal{S}$. Show that if X, Y are independent, then $S(\mathcal{A}_{X,Y}) = S(\mathcal{A}_X) + S(\mathcal{A}_Y)$.

CHEBYCHEV

Problem 5.4: You own an insurance company that gets random claims at random times. In order to have enough reserves, you want to estimate how large claims will be in the future. Your staff tells you the mean and standard deviation of the historical claim distribution but you do not know the distribution.
a) Why does Chebyshev's inequality imply that at least 89 percent of future claims will be within three standard deviations away from the mean?
b) Build a similar rule of thumb to see that [ ] percent of future claims are within two standard deviations from the mean. Explain.
c) Fill in the box: 96 percent of future claims are within [ ] standard deviations from the mean. Explain.

Problem 5.5: A probability space and random variable X defines what one calls a **null hypothesis**, the assumption that an effect does not exist. Assume you measure $X = c$ and that c is larger than the expectation, then the P**-value** of this experiment is defined as $\mathrm{P}[X \geq c]$. If the P-value is < 0.05, one considers the result as **significant** and rejects the null-hypothesis. If the P-value is > 0.05, one fails to reject the null hypothesis.
a) Assume a hypothesis is that X is exponentially distributed. You measure $X = 2$. What is the p-value?
b) Estimate the p-value using Chebyshev's inequality.
c) Having a p-value smaller than 0.05 is considered the gold standard for "statistical significance". Discuss the following strategy: we repeat an experiment a couple of times until the P-value is smaller than 5 percent. You label the early runs as warm-up-test runs and publish the paper.
d) Is it true that if you make a measurement and see the P value is larger than 0.05 that the non-significance means that the effect does exist? Explain in an example.

[1]

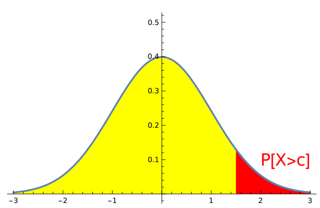

FIGURE 1. P-Value.

OLIVER KNILL, KNILL@MATH.HARVARD.EDU, MATH 154, SPRING, 2025

[1]If c was smaller than the expectation, we would define the P-value as $\mathrm{P}[X \leq c]$.