**Program of “Information theory and inference”**

1. Introduction: relation between information theory and inference. The “noisy channel” as a general setting for inference problems (*Lecture 1 slides*).
2. How to measure information for events. **Definition of Shannon entropy** from first principles (*Bialek §6.1, Allegra §1.1)*. Probability trees and information (*Allegra §1.1.2*).
3. **Conditional entropy** (*McKay §8.1, Cover §2.2, Allegra §1.4*), **Mutual information** (*McKay §8.1, Cover §2.3, Allegra §1.4*), **Kullback-Leibler divergence** (*McKay §2.6, Cover §2.3, Allegra §1.3*)
4. **Lossy compression** (*McKay §4.2,§4.3, Cover §3.2, Allegra §2.1*). **Typicality** (*McKay §4.4, Cover §3.1,§3.3, Allegra §2.1*). **Source coding theorem** (*McKay §4.3, Cover §3.2, Allegra §2.1.1*).
5. **Symbol codes** (*McKay §5.1, Cover §5.1, Allegra §2.2*). **Unique decodability** (*McKay §5.2, Cover §5.2, Allegra §2.1.1*). **Lossless Compression** (*McKay §5.3, §5.4, Cover §5.3, §5.4,§5.5, Allegra §2.2.3*). Huffman codes (*McKay §5.5, Cover §5.6, Allegra §2.2.4*).
6. **Noisy Channels** (*McKay §9.3, Cover §4.1, Cover §7.1,7.2*). Decoding over noisy channels and inference (*McKay §9.4, Allegra §3.1*).  **Channel capacity** (*McKay §9.5, Cover §7.1, §7.3*). Error-correcting codes (*McKay §1.2*).
7. Joint typicality (*McKay §10.2, Cover §7.6*). **The Noisy-Channel Coding Theorem** (*McKay §9.6,§10.1,§10.1,§10.3, Cover §7.4, §7.7, Allegra §3.3*).
8. **Differential entropy** (*Cover §8.1, §8.3,§8.4,§8.5, Allegra §4.1*), Gaussian channel (*McKay §11.1,§11.2,§11.3, Cover §9.1, §9.2, Allegra §4.2*).
9. Markov chains and entropy production (*Cover §4.1,§4.2, Allegra §5*).
10. Decision theory. Bayesian and max-likelihood estimators (*Barbier §2*).
11. The Monte Carlo idea, Monte Carlo Integration (*Campesan §2.1, Allegra §7.1*).
12. **Elementary sampling methods** (*Bishop §11.1, Campesan §1.2, Allegra §7.2*), **rejection sampling** (*McKay §29.3, Bishop §11.1*)
13. **Metropolis algorithm** (*McKay §29.4, Bishop §11.2, Campesan §2.2, Allegra §7.4*) and its convergence properties (*Campesan §2.1, Allegra §7.4.1*). Metropolis algorithm tuning (*Campesan §2.3.1,§2.3.2*)
14. **Gibbs sampling** (*McKay §29.5, Bishop §11.3,Campesan §2.4. Allegra §7.5*).
15. Clustering as an inference problem, latent variables (*Allegra §7.5.1*).
16. **Variational Bayes** (*Bishop §10, Allegra §8.1*). Variational autoencoders (*Allegra §8.3)*.
17. **Hamiltonian Monte Carlo** (*Brooks §5,Campesan §3.1, Allegra §9*). STAN.
18. **Approximate Bayes Computation** (*Campesan §4.1, Allegra §10*).

**References**

* Allegra, M., *Notes on information theory and inference,* Lecture Notes 2022/2023
* Bialek, W. (2012). *Biophysics: searching for principles*. Princeton University Press.
* Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine  
  learning* (Vol. 4, No. 4, p. 738). New York: springer.
* Brooks, S., Gelman, A., Jones, G., & Meng, X. L. (Eds.). (2011). *Handbook  
  of markov chain monte carlo*. CRC press
* Cover, T. M. , & Thomas, J. M. (1999). *Elements of information theory.* John Wiley & Sons.
* MacKay, D. J., & Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.
* Campesan G., Conforto F., Faorlin T., Marcomini A., Rosset L., Zanola A., *Information theory and inference,* Lecture Notes 2021/2022