The two text books you should know about are:

1. Cover, T. M., & Thomas, J. A. (2012). Elements of Information Theory. John Wiley & Sons.
2. MacKay, D. J. (2003). Information theory, inference and learning algorithms. Cambridge university press.

The MacKay book especially has many great examples and exercises, like the weighing one we did in class.

The neuroscience papers I mentioned are:

1. Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. Zeitschrift für Naturforschung c, 36(9-10), 910-912.
2. Friedrich, R. W., & Laurent, G. (2001). Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. Science, 291(5505), 889-894.

Here are several other papers on related topics:

1. Another way of implementing non-redundant representations, this time incorporating more biological constraints: Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381(6583), 607-609.
2. A more modern example of efficient coding being applied in neuroscience (and a very nice paper!): Młynarski, W. F., & Hermundstad, A. M. (2021). Efficient and adaptive sensory codes. Nature Neuroscience, 24(7), 998-1009.
3. An earlier paper in the same direction as above: Młynarski, W. F., & Hermundstad, A. M. (2018). Adaptive coding for dynamic sensory inference. Elife, 7, e32055.
4. A beautiful paper on reliable coding in grid cells (we didn't get to this part of the lecture, this is more related to channel coding rather than source coding): Sreenivasan, S., & Fiete, I. (2011). Grid cells generate an analog error-correcting code for singularly precise neural computation. Nature neuroscience, 14(10), 1330-1337.