Information Theory Notes

Charles Zheng and Yuval Benjamini

November 18, 2015

These are preliminary notes.

1 AEP for gaussian

Consider an infinite sequence of paired random vectors X^n, Y^n for $n = 1, 2, \ldots$, where for each n, (X^n, Y^n) is jointly multivariate gaussian with mean zero and covariance

$$\text{Cov}\begin{pmatrix} X^n \\ Y^n \end{pmatrix} = \Sigma^n = \begin{pmatrix} \Sigma_X^n & \Sigma_{XY}^n \\ \Sigma_{YX}^n & \Sigma_Y^n \end{pmatrix}.$$

Recall the following formulas for entropy (in bits):

$$H(X^n) = \frac{1}{2}\log_2(2\pi|e\Sigma_X^n|)$$

$$H(Y^n) = \frac{1}{2}\log_2(2\pi|e\Sigma_X^n|)$$

$$H(X^n, Y^n) = \frac{1}{2}\log_2(2\pi|e\Sigma^n|)$$

and for the log-2 densities:

$$\log p(x^n) = -\frac{1}{2}\log_2(2\pi|\Sigma_X^n|) - \frac{\log_2 e}{2}(x^n)^T(\Sigma_X^n)^{-1}(x^n)$$

$$\log p(y^n) = -\frac{1}{2}\log_2(2\pi|\Sigma_Y^n|) - \frac{\log_2 e}{2}(y^n)^T(\Sigma_Y^n)^{-1}(y^n)$$

$$\log p(x^n, y^n) = -\frac{1}{2}\log_2(2\pi|\Sigma^n|) - \frac{\log_2 e}{2}(x^n, y^n)^T(\Sigma^n)^{-1}(x^n, y^n)$$

For given n, define the set of *jointly typical values* $A_\epsilon^{(n)}$ as the set of pairs (x^n, y^n) such that

$$|-\log p(x^n) - H(X^n)| < \epsilon$$

$$|-\log p(y^n) - H(Y^n)| < \epsilon$$

$$|-\log p(x^n, y^n) - H(X^n, Y^n)| < \epsilon.$$

The standard joint AEP theorem for continuous random variables (Cover and Thomas 2006) yields the following result as a special case:

Corollary. *Suppose*

$$\Sigma_X^n = \begin{pmatrix} \Sigma_X^1 & 0 & \dots & 0 \\ 0 & \Sigma_X^1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Sigma_X^1 \end{pmatrix}$$

$$\Sigma_Y^n = \begin{pmatrix} \Sigma_Y^1 & 0 & \dots & 0 \\ 0 & \Sigma_Y^1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Sigma_Y^1 \end{pmatrix}$$

and

$$\Sigma_{XY}^n = \begin{pmatrix} \Sigma_{XY}^1 & 0 & \dots & 0 \\ 0 & \Sigma_{XY}^1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Sigma_{XY}^1 \end{pmatrix}.$$

Then:

- $\Pr[(X^n, Y^n) \in A_\epsilon^{(n)}] \to 1$ *as* $n \to \infty$.
- *Vol*$(A_\epsilon^{(n)}) \leq 2^{nH(X^1,Y^1)+\epsilon}$ *for all* n.
- *If* $(\tilde{X}^n, \tilde{Y}^n)$ *are multivariate normal with covariance* $\begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{pmatrix}$ *then*

  $$\Pr[(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}] \leq 2^{-n(I(X^1;Y^1)-3\epsilon)}$$

  Also, for sufficiently large n,

  $$\Pr[(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}] \geq (1-\epsilon)2^{-n(I(X;Y)+3\epsilon)}.$$

We are interested generalizing the above Corollary to a broader class of sequences $\Sigma^1, \Sigma^2, \ldots$.

It is sufficient if

$$x^T \Sigma_X^{-1} x \to \mathbf{E}[x^T \Sigma_X^{-1} x]$$
$$y^T \Sigma_Y^{-1} y \to \mathbf{E}[y^T \Sigma_Y^{-1} y]$$
$$(x, y)^T \Sigma^{-1} (x, y) \to \mathbf{E}[(x, y)^T \Sigma^{-1} (x, y)]$$

2 Fano's inequality

Suppose we draw $\mu_1, \ldots, \mu_K \sim N(0, I)$, and then for $i^* \sim Unif\{1, \ldots, K\}$ we draw $y^* \sim N(\mu_{i^*}, \Omega)$. We predict $\hat{i}$ using the Bayes' rule. Letting $p = \Pr[\hat{i} \neq i]$, then by Fano's inequality,

$$\log_2 K \leq \frac{H(p) + \frac{1}{2} \log \frac{|I+\Omega|}{|\Omega|}}{1 - p}.$$

where

$$H(p) = p \log_2 p + (1 - p) \log_2 (1 - p) \leq 1.$$

In particular, for $p = 1/2$,

$$\log_2 K \leq 2 + \log \frac{|I + \Omega|}{|\Omega|}.$$