Seminary 2 - Inequalities

Radu Trîmbiţaş

October 14, 2013

Problem 1 (Finite entropy) *Show that for a discrete random variable* $X \in \{1, 2, ...\}$*, if* $E \log X < \infty$*, then* $H(X) < \infty$*.*

Solution. Let the distribution on the integers be $p1, p2, \ldots,$. Then $H(p) = -\sum p_i \log p_i$ and $E \log X = \sum p_i \log i = c < 1$. We will now find the maximum entropy distribution subject to the constraint on the expected logarithm. Using Lagrange multipliers, we have the following functional to optimize

$$J(p) = -\sum p_i \log p_i - \lambda_1 \sum p_i - \lambda_2 \sum p_i \log i$$

Diferentiating with respect to p_i and setting to zero, we find that the p_i that maximizes the entropy set $p_i = ai^\lambda$, where $a = 1/\sum i^\lambda$ and chosed to meet the expected log constraint, i.e.

$$\sum i^\lambda \log i = c \sum i^\lambda.$$

Using this value of p_i, we can see that the entropy is finite. ■

Problem 2 (Example of joint entropy) *Let* $p(x, y)$ *be given by*

$X\backslash Y$	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

. Find

(a) $H(X), H(Y)$;

(b) $H(X|Y), H(Y|X)$;

(c) $H(X, Y)$;

(d) $H(Y) - H(Y|X)$;

(e) $I(X, Y)$.

(f) Draw a Venn diagram for the quantities in (a) through (e).

Solution.

(a) $H(X) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.918\,30 = H(Y)$

(b) $H(X|Y) = \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) = \frac{2}{3} \cdot \left(-2 \cdot \frac{1}{2}\log_2 \frac{1}{2}\right) = 0.666\,67 = H(Y|X)$.

(c) $H(X,Y) = -3 \cdot \frac{1}{3} \cdot \log_2 \frac{1}{3} = 1.585\,0$

(d) $H(Y) - H(Y|X) = 0.918\,30 - 0.666\,67 = 0.251\,63$

(e) $I(X,Y) = H(X) - H(Y|X) = 0.251\,63$.

■

Problem 3 (Entropy of a sum) *Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectively. Let $Z = X + Y$.*

(a) Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus the addition of independent random variables adds uncertainty.

(b) Give an example of (necessarily dependent) random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.

(c) Under what conditions does $H(Z) = H(X) + H(Y)$?

Solution.

(a) $Z = X + Y$. Hence $p(Z = z|X = x) = p(Y = z - x|X = x)$.

$$\begin{aligned} H(Z|X) &= \sum p(x) p(Z|X=x) \\ &= -\sum_x p(x) \sum_z p(Z=z|X=x) \log p(Z=z|X=x) \\ &= \sum_x p(x) \sum_z p(Y=z-x|X=x) \log p(Y=z-x|X=x) \\ &= \sum_x p(x) p(Y|X) = H(Y|X). \end{aligned}$$

If X and Y are independent, then $H(Y|X) = H(Y)$. Since $I(X;Z) \geq 0$, we have $H(Z) \geq H(Z|X) = H(Y|X) = H(Y)$. Similarly we can show that $H(Z) \geq H(X)$.

(b) Consider the following joint distribution for X and Y Let

$$X = -Y = \begin{cases} 1 & \text{with prob } 1/2 \\ 0 & \text{with prob } 1/2 \end{cases}$$

$H(X) = H(Y) = 1$, but $Z = 0$ with probability 1 and $H(Z) = 0$.

(c) We have

$$H(Z) \leq H(X,Y) \leq H(X) + H(Y)$$

because Z is a function of (X,Y) and $H(X,Y) = H(X) + H(Y|X) H(X) + H(Y)$. We have equality iff (X,Y) is a function of Z and $H(Y) = H(Y|X)$, i.e., X and Y are independent.

■

Problem 4 (Data processing) . *Let $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ form a Markov chain in this order; i.e., let*

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_n|x_{n-1}).$$

Reduce $I(X_1, X_2, \ldots, X_n)$ to its simplest form.

Solution. By the chain rule for mutual information,

$$I(X_1, X_2, \ldots, X_n) = I(X_1, X_2) + I(X_1, X_3|X_2) + +I(X_1, X_n|X_2, \ldots, X_{n-2}).$$

By the Markov property, the past and the future are conditionally independent given the present and hence all terms except the first are zero. Therefore

$$I(X_1, X_2, \ldots, X_n) = I(X_1, X_2).$$

■

Problem 5 (Bottleneck) *Suppose a (non-stationary) Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus $X_1 \to X_2 \to X_3$, i.e., $p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$, for all $x_1 \in \{1, 2, \ldots, n\}$, $x_2 \in \{1, 2, \ldots, k\}$, $x_3 \in \{1, 2, \ldots, m\}$.*

(a) *Show that the dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X1, X3) \leq \log k$.*

(b) *Evaluate $I(X1; X3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.*

Solution.

(a) From the data processing inequality, and the fact that entropy is maximum for a uniform distribution, we get

$$\begin{aligned} I(X_1; X_3) &\leq I(X_1; X_2) \\ &= H(X_2) - H(X_2|X_1) \\ &\leq H(X_2) \leq \log k. \end{aligned}$$

Thus, the dependence between X_1 and X_3 is limited by the size of the bottleneck. That is $I(X_1; X_3) \leq \log k$.

(b) For$k = 1$, $I(X1; X3) \leq \log 1 = 0$ and since $I(X1; X3) \geq 0$, $I(X1; X3) = 0$. Thus, for $k = 1$, X_1 and X_3 are independent.

■

Problem 6 (Infinite entropy) *This problem shows that the entropy of a discrete random variable can be infinite. Let* $A = \sum_{n=2}^{\infty}(nlog^2 n)^{-1}$. *(It is easy to show that* A *is finite by bounding the infinite sum by the integral of* $(x\log^2 x)^{-1}$.*) Show that the integer-valued random variable* X *defined by* $P(X = n) = (Anlog^2 n)^{-1}$ *for* $n = 2, 3, \dots,$ *has* $H(X) = +\infty$.

Solution. By definition, $p_n = Pr(X = n) = 1/An\log^2 n$ for $n \geq 2$. Therefore

$$\begin{aligned}
H(X) &= -\sum_{n=2}^{\infty} p(n)\log p(n) \\
&= -\sum_{n=2}^{\infty} \left(1/An\log^2 n\right)\log\left(1/An\log^2 n\right) \\
&= \sum_{n=2}^{\infty} \frac{\log\left(1/An\log^2 n\right)}{An\log^2 n} \\
&= \sum_{n=2}^{\infty} \frac{\log A + \log n + 2\log\log n}{An\log^2 n} \\
&= \log A + \sum_{n=2}^{\infty}\frac{1}{An\log n} + \sum_{n=2}^{\infty}\frac{2\log\log n}{An\log^2 n}.
\end{aligned}$$

The first term is finite. For base 2 logarithms, all the elements in the sum in the last term are nonnegative. (For any other base, the terms of the last sum eventually all become positive.) So all we have to do is bound the middle sum, which we do by comparing with an integral.

$$\sum_{n=2}^{\infty}\frac{1}{An\log n} > \int_{2}^{\infty}\frac{1}{Ax\log x}\mathrm{d}\,x = K\ln\ln x\big|_{x=2}^{\infty} = +\infty.$$

Thus $H(X) = +\infty$. ∎

Problem 7 (Run length coding) *Let* $X_1, X_2, \dots, X_n$ *be (possibly dependent) binary random variables. Suppose one calculates the run lengths* $R = (R_1, R_2, \dots)$ *of this sequence (in order as they occur). For example, the sequence* $X =$ 0001100100 *yields run lengths* $\mathbf{R} = (3, 2, 2, 1, 2)$. *Compare* $H(X_1, X_2, \dots, X_n)$, $H(\mathbf{R})$ *and* $H(X_n, \mathbf{R})$. *Show all equalities and inequalities, and bound all the differences.*

Solution. Since the run lengths are a function of $X_1, X_2, \dots, X_n$, $H(\mathbf{R}) \leq H(X)$. Any X_i together with the run lengths determine the entire sequence $X_1, X_2, \dots, X_n$. Hence

$$\begin{aligned}
H(X_1, X_2, \dots, X_n) &= H(X_i, \mathbf{R}) \\
&= H(\mathbf{R}) + H(Xi|\mathbf{R}) \\
&\leq H(\mathbf{R}) + H(X_i) \\
&\leq H(\mathbf{R}) + 1.
\end{aligned}$$

∎

Problem 8 *Let $p(x)$ be a probability mass function. Prove, for all $d \geq 0$,*

$$P\left(p(X) \leq d\right) \log \frac{1}{d} \leq H(X).$$

Solution.

$$\begin{aligned}
P\left(p(X) \leq d\right) \log \frac{1}{d} &= \sum_{x: p(x) < d} p(x) \log \frac{1}{d} \\
&\leq \sum_{x: p(x) < d} p(x) \log \frac{1}{p(x)} \\
&\leq \sum_{x} p(x) \log \frac{1}{p(x)} \\
&= H(X).
\end{aligned}$$

■

Problem 9 (Maximum entropy) *Find the probability mass function $p(x)$ that maximizes the entropy $H(X)$ of a non-negative integer-valued random variable X subject to the constraint*

$$EX = \sum_{n=0}^{\infty} n p(n) = A,$$

for a fixed value $A > 0$. Evaluate this maximum $H(X)$.

Solution. We must minimize

$$H(X) = -\sum_{n=0}^{\infty} p(n) \log_2 p(n),$$

subject to the constraints

$$\begin{aligned}
\sum_{n=0}^{\infty} p(n) &= 1 \\
\sum_{n=0}^{\infty} n p(n) &= A.
\end{aligned}$$

Lagrange multipliers: we have to find stationary points of

$$H(p_n, \lambda_1, \lambda_2) = -\sum_{n=0}^{\infty} p_n \log_2 p_n + \lambda_1 \left(\sum_{n=0}^{\infty} p_n - 1 \right) + \lambda_2 \left(\sum_{n=0}^{\infty} n p_n - A \right)$$

$$\begin{aligned}
\frac{\partial H}{\partial p_n} &= -\log_2 p_n - p_n \frac{1}{p_n \ln 2} + \lambda_1 + n\lambda_2 \\
&= -\log_2 p_n - \frac{1}{\ln 2} + \lambda_1 + n\lambda_2 = 0
\end{aligned}$$

Solution is:

$$p_n = \frac{2^{\lambda_1 + n\lambda_2}}{e} = \alpha\beta^n.$$

$-H(p)$ convex$\Longrightarrow$ $H(p)$ concave, and it has only one maximum.

From our constraints it follows

$$\sum_{n=0}^{\infty} \alpha\beta^n = \frac{\alpha}{1-\beta} = 1$$
$$\sum_{n=0}^{\infty} n\alpha\beta^n = \frac{\alpha\beta}{(1-\beta)^2} = A.$$

This allows us to obtain α and β

$$\frac{\alpha}{1-\beta} = 1$$
$$\frac{\alpha\beta}{(1-\beta)^2} = A$$

Solution is

$$\begin{cases} \alpha = \frac{1}{A+1}, \\ \beta = \frac{A}{A+1} \end{cases}$$

so,

$$p_n = \frac{1}{A+1}\left(\frac{A}{A+1}\right)^n$$

The maximum entropy is

$$H(X) = (A+1)\log(A+1) - A\log A$$

■

Problem 10 (Fano) *We are given the following joint distribution on (X, Y)*

$X \backslash Y$	a	b	c
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

. *Let $\widehat{X}(Y)$ be an estimator for X (based on Y) and let $P_e = P\{\widehat{X}(Y) \neq X\}$.*

(a) Find the minimum probability of error estimator $\widehat{X}(Y)$ and the associated P_e.

(b) Evaluate Fano's inequality for this problem and compare.

Solution.

(a) From inspection we see that

$$\widehat{X}(Y) = \begin{cases} 1 & y = a \\ 2 & y = b \\ 3 & y = c \end{cases}.$$

For probability of error

$$\begin{aligned} P_e &= P(1,b) + P(1,c) + P(2,a) + P(2,c) + P(3,a) + P(3,b) \\ &= \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{1}{2}. \end{aligned}$$

(b) From Fano's inequality we know

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|}$$

$$\begin{aligned} H(X|Y) &= P(Y=a)H(X|Y=a) + P(Y=b)H(X|Y=b) + P(Y=c)H(X|Y=c) \\ &= 3\frac{1}{3}H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = \frac{3}{2}. \end{aligned}$$

Hence

$$P_e \geq \frac{\frac{3}{2} - 1}{\log_2 3} = 0.315\,46.$$

Hence our estimator $\widehat{X}(Y)$ is not very close to Fano's bound in this form. If $\widehat{X} \in \mathcal{X}$; as it does here, we can use the stronger form of Fano's inequality to get

$$P_e \geq \frac{H(X|Y) - 1}{\log (|\mathcal{X}| - 1)} = \frac{\frac{3}{2} - 1}{\log_2 2} = \frac{1}{2}.$$

Therefore our estimator $\widehat{X}(Y)$ is actually quite good.

■

Problem 11 (Fano's inequality) *Let $P(X = i) = p_i$, $i = 1, 2, \ldots, m$ and let $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_m$. The minimal probability of error predictor of X is $\widehat{X} = 1$, with resulting probability of error $P_e = 1 - p_1$. Maximize $H(\mathbf{p})$ subject to the constraint $1 - p_1 = P_e$ to find a bound on P_e in terms of H. This is Fano's inequality in the absence of conditioning.*

Solution. The minimal probability of error predictor when there is no information is $\widehat{X} = 1$, the most probable value of X. The probability of error in this case is $P_e = 1 - p_1$. Hence if we fix P_e, we fix p_1. We maximize the entropy of

X for a given P_e to obtain an upper bound on the entropy for a given P_e. The entropy,

$$\begin{aligned}
H(\mathbf{p}) &= -p_1 \log p_1 - \sum_{i=2}^{m} p_i \log p_i \\
&= -p_1 \log p_1 - \sum_{i=2}^{m} P_e \frac{p_i}{P_e} \log \frac{p_i}{P_e} - P_e \log P_e \\
&= H(P_e) + P_e H\left(\frac{p_2}{P_e}, \dots \frac{p_m}{P_e}\right) \\
&\leq H(P_e) + P_e \log(m-1)
\end{aligned}$$

since the maximum of $H\left(\frac{p_2}{P_e}, \dots \frac{p_m}{P_e}\right)$ is attained by an uniform distribution. Hence any X that can be predicted with a probability of error P_e must satisfy

$$H(X) \leq H(P_e) + P_e \log(m-1),$$

which is the unconditional form of Fano's inequality. We can weaken this inequality to obtain an explicit lower bound for P_e,

$$P_e \geq \frac{H(X)-1}{\log(m-1)}.$$

■