Data Science. Group Assignment 3

Ekaterina Toropova, Marina Dushkina, Haijia Chen,
Tse-An Hsu, Siyuan Lai

December 2022

1 Exercise 5.20

Show that the relative entropy risk is the same risk as we saw in Section 4.2, it only differs by a constant.

Proof: Let F and G be any two distribution functions with density f and g belonging to a model space M,let $X = (X_1, ...X_n)$ be i.i.d. from distribution function F. According to section 5.3,the relative entropy risk is

$$R(G) = \int \ln \frac{f(x)}{g(x)} f(x)dx = -\int \ln \frac{g(x)}{f(x)} f(x)dx$$

then, according to the definition of expected value,the above could be written as

$$R(X) = \mathbb{E}[-\ln \frac{f(x)}{g(x)}]$$

Given X is a i.i.d random variable,then the empirical Risk becomes

$$R(X) = -\frac{1}{n} \sum (\ln \frac{f(X_i)}{g(X_i)})$$

Let P be another distribution function with density p,according to 4.2, the risk is defined as

$$R^*(X) = -\int \ln(p(x))p(x)dx$$

As above,the empirical Risk becomes

$$R^*(X) = -\frac{1}{n} \sum \ln(P(X_i))$$

As $P(X_i)$ and $f(X_i)/g(X_i)$ are different constants , so proof ends.

2 Lemma 6.8

Consider a congruential generator D on $\mathcal{M} = \{0, 1, ..., M-1\}$ with period M, then for any starting point $u_0 \in M$, the sequence $u_i = D(u_{i-1})$ is pseudorandom on M.

Proof:

Since the generator has period M, each element of $\mathcal{M}$ occurs exactly once in the sub sequence $u_1, ..., u_M$. Otherwise, if an element $m \in \mathcal{M}$ occured twice, i.e. $u_i = m$ and $u_j = m$ with $1 \leq i, j \leq M$, then $u_{i+1} = u_{j+1}, u_{i+2} = u_{j+2}$ and so on, and the period in this case is $|j - (i-1)| < M$. The same holds for every sub sequence $u_{Mk+1}, ..., u_{M(k+1)}$. Hence, we can write the following estimate for every $m \in \mathcal{M}, n \in \mathbb{N}$ and $k = \left[\frac{n}{M}\right]$

$$k \leq N_n(m) \leq k+1,$$

where $N_n(m)$ denotes the number of $u_i = m$ for $i = 1, ..., n$. Dividing this by n, we get

$$\frac{k}{n} \leq \frac{N_n(m)}{n} \leq \frac{k+1}{n}.$$

Applying $kM \leq n \leq (k+1)M$, it is easy to see that

$$\frac{k}{M(k+1)} \leq \frac{N_n(m)}{n} \leq \frac{k+1}{kM}.$$

When n goes to infinity k also goes to infinity, hence $\dfrac{k}{k+1} \to 1$ and $\dfrac{k+1}{k} \to 1$. Then by the sandwich theorem $\dfrac{N_n(m)}{n} \to \dfrac{1}{M}$.

3 Exercise 6.19

Suppose $X \sim N(0,1)$ and $Y \sim N(0,1)$, they are independent random variables, their PDFs are:

$$p(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}, p(Y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Y^2}{2}}$$

The joint PDF is:

$$p(X,Y) = p(X) \cdot p(Y) = \frac{1}{2\pi} e^{-\frac{X^2+Y^2}{2}}$$

Convert X, Y into polar coordinates:

$$X = R\cos\theta, Y = R\sin\theta$$

Then

$$\frac{1}{2\pi} e^{-\frac{X^2+Y^2}{2}} = \frac{1}{2\pi} e^{-\frac{R^2}{2}}$$

Therefore,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{X^2+Y^2}{2}} dX dY = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{R^2}{2}} R d\theta dR = 1$$

Consequently, the distribution functions are:

$$P_R(R \leq r) = \int_0^{2\pi} \int_0^r \frac{1}{2\pi} e^{-\frac{R^2}{2}} R d\theta dR = 1 - e^{-\frac{r^2}{2}}$$
$$P_\theta(\theta \leq \phi) = \int_0^{\phi} \int_0^{\infty} \frac{1}{2\pi} e^{-\frac{R^2}{2}} R d\theta dR = \frac{\phi}{2\pi}$$

Notice $\theta \sim U(0, 2\pi)$, recall the inversion sampling method from Theorem 5.38, the inverse function of $P(R \leq r)$ is

$$R = F_R^{-1}(z) = \sqrt{-2\ln(1-z)}$$

When $Z \sim U(0,1)$, the distribution function of R is $P(R \leq r)$, choose $U_1, U_2 \overset{\text{i.i.d}}{\sim} U(0,1)$, thus

$$\theta = 2\pi U_2, 1 - z = U_1, R = \sqrt{-2\ln U_1}$$

Substitute them into $X = R\cos\theta, Y = R\sin\theta$, we have

$$X = \sqrt{-2\ln\left(U_1\right)}\cos\left(2\pi U_2\right)$$
$$Y = \sqrt{-2\ln\left(U_1\right)}\sin\left(2\pi U_2\right)$$

4 Exercise 7.12

Prove Lemma 7.11 in a similar way to Lemma 7.7.
Proof: Let us start by applying the law of total probability

$$P(X_t = s_i) = \sum_{x_{t-1}} P(X_t = s_i | X_{t-1} = x_{t-1}) P(X_{t-1} = x_{t-1})$$
$$= \sum_{x_{t-1}} P_t(x_{t-1}, s_i) P(X_{t-1} = x_{t-1})$$

Since n was arbitrary we can apply it again until we reach X_0, namely

$$P(X_t = s_i) = \mu_t(s_i) = \sum_{x_0, x_1, \ldots, x_{t-1}} P_t(x_{t-1}, s_i) P_{t-1}(x_{t-2}, x_{t-1}) \ldots P_1(x_0, x_1) P(X_0 = x_0)$$

The above is just an element inside μ_t, we could do the same calculation on other states. If we combine all the elements of μ_t, then we get

$$\mu_t = \mu_0 P_1 P_2 P_3 \ldots . P_t$$

5 Exercise 7.17

Theorem 7.16 Let $W_1, ..., \overset{iid}{\sim} F$ such that (ρ_t, W_t) is a random mapping representation (RMR) for a transition matrix P_t, for all $t \in \mathbb{N}$. Then if $X_0 \sim \mu_0$,

$$X_t := \rho_t(X_{t-1}, W_t), t \in \mathbb{N},$$

is a Markov chain with initial distribution μ_0 and transition matrix P_t at t.

Proof:

To prove the theorem, we first need to show that the sequence $X_0, X_1, X_2, \ldots$ of random variables satisfies the Markov property.

The Markov property states that, given the present state of a system, its future and past are independent. In other words, if X_t is a Markov chain with transition matrix P_t, then for all $t \geq 0$:

$$P[X_{t+1} = x | X_0, X_1, \ldots, X_t] = P[X_{t+1} = x | X_t]$$

To do this, we can use the definition of the sequence X_t as a function of the previous state X_{t-1} and the random variable W_t. We have:

$$P[X_{t+1} = x | X_0, X_1, \ldots, X_t] = P[\rho_{t+1}(X_t, W_{t+1}) = x | X_0, X_1, \ldots, X_t] =$$

$$= P[\rho_{t+1}(X_t, W_{t+1}) = x | X_t].$$

The last equality holds since X_{t+1} is fully determined by X_t and doesn't depend on $X_0, X_1, \ldots, X_t$.

Secondly, we need to prove that:

$$P(X_{t+1} = y | X_t = x) = P_{t+1}(x, y).$$

We prove it as follows: $P(X_{t+1} = y | X_t = x) = P(\rho_{t+1}(X_t, W_{t+1} = y | X_t = x) =$

$$= P(\rho_{t+1}(x, W_{t+1}) = y) = P_{t+1}(x, y).$$