Intro DS - Group Assignment 3

Albin Åberg Dahlberg, Emil Wormbs,
Folke Hilding, Oscar Boman, Sofia Alfsson

December 27, 2023

Excersize 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

Answer: For estimation on non-parametric DF, the underlying DF and model is F^* with density f^*, where $F^* \in \mathcal{E}, F^* \in \mathcal{M}_0$ and $\mathcal{E}$ is the statistical model and $\mathcal{M}_0$ is the model space.
For any distribution $G \in \mathcal{M}_0$ with density g, its *relative entropy risk* was defined as

$$R(G) = \int \ln\left(\frac{f^*(x)}{g(x)}\right) f^*(x)dx.$$

Relative entropy risk can be decomposed

$$R(G) = \int \ln\left(\frac{f^*(x)}{g(x)}\right) f^*(x)dx \tag{1}$$

$$= \int \ln(f^*(x)) - \ln(g(x)) f^*(x)dx \tag{2}$$

$$= \int \ln(f^*(x)) f^*(x)dx - \int \ln(g(x)) f^*(x)dx \tag{3}$$

Looking at section 4.2, we assume that α^* is our underlying model parameters with density p_{α^*}. The risk of the model with parameter α, with some density p_α is then defined as

$$R(\alpha) = -\int \ln(p_\alpha(z)) p_{\alpha^*}(z)dz \tag{4}$$

We know that f^* and p_{α^*} represent the underlying model, and as such $f^* = p_{\alpha^*}$. The same goes for $g = p_\alpha$, the density of any distribution. Looking at (1), the first term is the negative risk of f^*, where the underlying model is f^*. This risk is 0 plus some constant C due to primitive functions. As such, the *relative* *entropy risk* in (1) can be rewritten into

$$\begin{aligned} R(G) &= \int \ln\left(f^*(x)\right) f^*(x)dx - \int \ln\left(g(x)\right) f^*(x)dx \\ &= 0 + C - \int \ln\left(g(x)\right) f^*(x)dx \\ &= C - \int \ln(p_\alpha(x)) p_{\alpha^*}(x)dx \\ &= C + R(\alpha), \end{aligned} \tag{5}$$

that is, the risk from Section 4.2, plus some constant. □

Solve Exercise 6.11

Question: Prove lemma 6.10, Consider a congruential generator D on $\theta = \{0, 1, ..., M-1\}$ with period M, then for any starting point $u_0 \in \theta$, define $u_i = D(u_{i-1})$ then the sequence $v_i = u_i$ mod K for $1 \leq K \leq M$ is pseudorandom on $\{0, 1, ..., K-1\}$ if M is a multiple of K.

First of, as M is a multiple of K, such that $1 \leq K \leq M$, we can introduce a scalar $\lambda \in (0, 1]$ such that $K = \lambda M$ and $\{0, 1, ..., \lambda M - 1\}$. We also introduce a new congruential generator such that $v_i = D_v(v_{i-1})$ and $D_v(x) = u_i \bmod \lambda M$

$$D_v(D(x)) = ((ax + b) \bmod M) \bmod \lambda M \tag{6}$$

since $1 \leq \lambda M \leq M$ we can simplify this;

$$D_v(x) = (ax + b) \bmod \lambda M \tag{7}$$

$v_i = D_v(u_{i-1})$ holds for any starting point $u_0 \in \theta$, and as per lemma 6.8, since $D_v(x)$ is a congruential generator on $\theta_v = \{0, 1, ..., \lambda M - 1\}$ with period λM, the sequence is **psudorandom**.

Solve Exercise 6.19

First step is to prove that W, Y are independent. The functions for W, Y are:

$$\begin{aligned} Y &= |Z|^2 \\ W &= \frac{Z}{|Z|} \end{aligned}$$

Since both W, Y only depend on Z and none of them depend on each other, then their joint probability can be expressed as a product of their marginal probabilities:

$$P(W = w, Y = y) = P(W = w) \cdot P(Y = y)$$

Which means they are independent.

Next step is to show that if W is generated using $(\cos(2\pi U_2), \sin(2\pi U_2))$ then it is uniform on the unit circle. We know that U_2 is uniformly distributed on $[0, 1]$. Given the trigonometric identity:

$$\cos^2(\theta) + \sin^2(\theta) = 1$$

which is a property of points on the unit circle. We can apply normalization to $(\cos(2\pi U_2), \sin(2\pi U_2))$ to get:

$$\cos^2(2\pi U_2) + \sin^2(2\pi U_2) = 1$$

which shows that W is on the unit circle. In order to show that it's uniform, U_2 is uniformly distributed on $[0, 1]$ which means that $2\pi U_2$ is uniform on $[0, 2\pi]$.

Last step is to show that Z_0, Z_1 are independent. Z_0, Z_1 are defined as:

$$Z_0 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2)$$
$$Z_1 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

In order to show their independence it is enough to prove that the covariance between them is zero:

$$\mathrm{cov}(Z_0, Z_1) = \mathbb{E}[(Z_0 - \mathbb{E}[Z_0])(Z_1 - \mathbb{E}[Z_1])] = 0$$

Which can be written as:

$$\mathbb{E}[Z_0 \cdot Z_1] - \mathbb{E}[Z_0] \cdot \mathbb{E}[Z_1] = 0$$

We can now calculate $\mathbb{E}[Z_0 \cdot Z_1]$ which is done by:

$$\mathbb{E}[Z_0 \cdot Z_1] = \int_0^1 \int_0^1 \sqrt{(-2\ln(U_1)}\cos(2\pi U_2)) \cdot \sqrt{(-2\ln(U_1)}\sin(2\pi U_2))\, dU_1\, dU_2 = 0.$$

Since $\cos(2\pi U_2)$ and $\sin(2\pi U_2)$ are orthogonal the expression above can be evaluated to zero meaning:

$$\mathbb{E}[Z_0 \cdot Z_1] = 0$$

which means that:

$$\begin{aligned} \mathbb{E}[Z_0 \cdot Z_1] - \mathbb{E}[Z_0] \cdot \mathbb{E}[Z_1] &= 0 \\ 0 - \mathbb{E}[Z_0] \cdot \mathbb{E}[Z_1] &= 0 \\ 0 &= -\mathbb{E}[Z_0] \cdot \mathbb{E}[Z_1] \\ \mathbb{E}[Z_0] \cdot \mathbb{E}[Z_1] &= 0 \end{aligned}$$

This proves that the covariance between Z_0, Z_1 is zero and since they are Gaussian it means they are independent.

Solve Exercise 7.12

Prove Lemma 7.11 in a similar way to Lemma 7.7

Lemma 7.11: For a finite inhomogeneous Markov chain$(X_t)_{t\in\mathbb{Z}_+}$ with state space $\mathbb{X} = \{s_1, s_2, \cdots, s_k\}$ and initial distribution

$$\mu_0 := (\mu_0(s_1), \mu_0(s_2), \cdots, \mu_0(s_k))$$

where $\mu_0(s_i) = \mathbb{P}(X_0 = s_i)$, and transition matrices

$$(P_1, P_2, \cdots), P_t := (P_t(s_i, s_j))_{(s_i,s_j)\in\mathbb{X}\times\mathbb{X}}, t \in \{1, 2, \cdots\}$$

we have for any $t \in \mathbb{Z}_+$ that the distribution at time t given by:

$$\mu_t := (\mu_t(s_1), \mu_t(s_2), \cdots, \mu_t(s_k))$$

where $\mu_t(s_i) = \mathbb{P}(X_t = s_i)$ satisfies:

$$\mu_t = \mu_0 P_1 P_2 \cdots P_t$$

Or in plain English, the state at time t is given by multiplying the initial state by the transition matrices, which are ordered, from 1 through t. We start by investigating what $\mathbb{P}(X_{t+1})$ depends on.

$$\mathbb{P}(X_{t+1} = s|X_0, X_1, \cdots, X_t) = \mathbb{P}(X_{t+1} = s|X_t)$$

Or, each distribution depends on the transition matrix of the previous step. We can then use this fact to apply the law of total probability to write the PMF, or μ_t, of the random variable X_n as

$$\begin{aligned}\mathbb{P}(X_n = s_n) &= \sum_{s_{n-1}} \mathbb{P}(X_n = s_n|X_{n-1} = s_{n-1})\mathbb{P}(X_{n-1} = s_{n-1}) \\ &= \sum_{s_{n-1}} P_{s_{n-1}s_n}\mathbb{P}(X_{n-1} = s_{n-1})\end{aligned}$$

And, like in Lemma 7.7, as this step was chosen arbitrarily we can repeat it until we reach X_0 which gives us. Also note that $P_{s_a s_b}$ denotes the transition matrix from a to b

$$\mathbb{P}(X_n = s_n) = \sum_{s_0,\cdots,s_{n-1}} P_{s_{n-1}s_n} \cdots P_{s_0 s_1}\mathbb{P}(X_0 = s_0)$$

We then use $\mu_t(s_i) = \mathbb{P}(X_t = s_i)$ and plug in the entire state space $\mathbb{X} = \{s_1, s_2, \cdots, s_k\}$

$$\begin{aligned}\mathbb{P}(X_n = s_n) &= \sum_{s_0,\cdots,s_{n-1}} P_{s_{n-1}s_n} \cdots P_{s_0 s_1}\mathbb{P}(X_{n-1} = s_{n-1}) \\ \mu_t &= \sum_{s_0,\cdots,s_{n-1}} P_{s_{n-1}s_n} \cdots P_{s_0 s_1}\mathbb{P}(X_{n-1} = s_{n-1}) \\ &= \mu_0 P_1 P_2 \cdots P_t\end{aligned}$$

Solve Exercise 7.17

Do the proof of Theorem 7.16 by using the necessary Definitions.

Theorem 7.16:

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

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

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

Proof

We know by definition that the initial distribution of the Markov Chain is μ_0 as $X_0 \sim \mu_0$ is given in the Theorem.

The transition matrix P_t is defined by (ρ_t, W_t) for all $t \in \mathbb{N}$, meaning that the transition $X_{t-1}$to X_t has the RMR (ρ_t, W_t).

In order to prove that X_t fulfills the conditions for a Markov chain, we look at the Markov Property, meaning that the future evolution in a stochastic process is independent of its history - it is memoryless.

For X_t we hence have the conditional probability

$\mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}, ..., X_0 = x_0)$
$= \mathbb{P}(p_t(X_{t-1}, W_t) = x_t | X_{t-1} = x_{t-1}, ..., X_0 = x_0)$
$= \mathbb{P}(p_t(x_{t-1}, W_t) = x_t) = \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1})$

Which shows that the process fulfills the Markov property of being memoryless - meaning that the probability that X_t only takes on a value x_t depending on the *current* state, X_{t-1}. Hence it is a Markov chain with initial distribution μ_0 and transition matrix P_t at time t.

$\square$