INTRODUCTION TO STATISTICS

CARSON JAMES

Contents

1.	Introduction	1
2.	Sampling	2
2.1.	. Inverse CDF Sampling	2
2.2.	. Importance Sampling	3
2.3.	. Rejection Sampling	4
3.	Decision Theory	6
3.1.	. Introduction	6
3.2.	. Bayes Risk	6
3.3.	. Minimax Estimation	7
4.	Posterior Consistency	8
4.1.	. Introduction	8

1. Introduction

Definition 1.0.1. Let $A \in \mathcal{B}(\mathbb{R}^d)$ and $\Theta \neq \emptyset$. Suppose that m(A) > 0. We define

$$\mathcal{D}(A) = \{ f \in L^1(A) : f \ge 0 \text{ and } ||f||_1 = 1 \}$$

and for $\theta \in \Theta$, we define

$$\mathcal{D}(A|\theta) = \{ f : A \times \Theta \to \mathbb{R} : f(\cdot|\theta) \in \mathcal{D}(A) \}$$

2. Sampling

2.1. Inverse CDF Sampling.

2.2. Importance Sampling.

2.3. Rejection Sampling.

Exercise 2.3.1. Let $f, g \in \mathcal{D}(\mathbb{R}^d)$ and $A \in \mathcal{B}(\mathbb{R}^d)$. Suppose that $m^d(A) > 0$. If $X \sim f$, then $X|X \in A \sim ||fI_A||_1^{-1} fI_A$.

Proof. Let $C \in \mathcal{B}(\mathbb{R}^d)$. Then

$$P(X \in C | X \in A) = P(X \in C \cap A)P(X \in A)^{-1}$$
$$= ||fI_A||_1^{-1} \int_C fI_A dm^d$$

So
$$f_{X|X\in A} = ||fI_A||_1^{-1} fI_A$$
.

Exercise 2.3.2. Let $A, B \in \mathcal{B}(\mathbb{R}^d)$. Suppose that $A \subset B$ and $0 < m^d(A)$ and $m^d(B) < \infty$. If $X \sim \mathrm{Uni}(B)$, then $X | X \in A \sim \mathrm{Uni}(A)$.

Proof. Clear using the previous exercise with $f = I_B$.

Exercise 2.3.3. (Fundamental Theorem of Simulation):

Let $f \in \mathcal{D}(\mathbb{R}^d)$ and c > 0. Define

$$G_c = \{(x, v) \in \mathbb{R}^{d+1} : 0 < v < cf(x)\}$$

- (1) If $X \sim f$ and $U \sim \text{Uni}(0,1)$ are independent, then $(X, cUf(X)) \sim \text{Uni}(G_c)$.
- (2) If $(X, V) \sim \text{Uni}(G_c)$, then $X \sim f$.

Proof. First we note that $m^{d+1}(G_c) = c$.

(1) Suppose that $X \sim f$ and $U \sim \text{Uni}(0,1)$ are independent and put Y = cUf(X). Then $Y|X = x \sim cUf(x) \sim \text{Uni}(0,cf(x))$ and we have that for each $x \in \text{supp } X$ and $y \in (0,cf(x))$,

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f(x)$$
$$= \frac{1}{cf(x)}f(x)$$
$$= \frac{1}{c}$$

So $(X, Y) \sim \text{Uni}(G_c)$

(2) Suppose that $(X, V) \sim \text{Uni}(G_c)$. Then $f_{X,V}(x, v) = \frac{1}{c}I_{G_c}(x, v)$. So

$$f_X(x) = \int_{\mathbb{R}} \frac{1}{c} I_{G_c}(x, v) dm(v)$$
$$= \int_0^{cf(x)} \frac{1}{c} dv$$
$$= f(x)$$

So $X \sim f$.

Exercise 2.3.4. Let $f, g \in \mathcal{D}(\mathbb{R}^d)$, $c_f, c_g > 0$ and M > 0. Put $\tilde{f} = c_f f$ and $\tilde{g} = c_g g$. Suppose that $\tilde{f} \leq M\tilde{g}$. If $Y \sim g$ and $U \sim \mathrm{Uni}(0,1)$ are independent, then $Y|U \leq \frac{\tilde{f}(Y)}{M\tilde{g}(Y)} \sim f$ and $P(U \leq \frac{\tilde{f}(Y)}{M\tilde{g}(Y)}) = \frac{c_f}{c_g M}$

Proof. Put

$$G_g = \{ (y, v) \in \mathbb{R}^{d+1} : 0 < v < M\tilde{g}(y) \}$$

and

$$G_f = \{(y, v) \in \mathbb{R}^{d+1} : 0 < v < \tilde{f}(y)\}$$

Then $G_f \subset G_g$, $m^d(G_g) = c_g M$ and $m^d(G_f) = c_f$. By the first part of the fundamental theorem of simulation, we know that

$$(Y, MUc_gg(Y)) \sim \text{Uni}(G_g)$$

Since $\{(Y, MUc_gg(Y)) \in G_f\} = \{U \leq \frac{c_ff(Y)}{Mc_gg(Y)}\}$, a previous exercise tells us that

$$(Y, MUc_gg(Y))|U \le \frac{c_ff(Y)}{Mc_gg(Y)} \sim \text{Uni}(G_f)$$

Then the second part of the fundamental theorem of simulation tells us that

$$Y|U \le \frac{c_f f(Y)}{M c_g g(Y)} \sim f$$

Finally we have that

$$P\left(U \le \frac{c_f f(Y)}{M c_g g(Y)}\right) = P[(Y, M U c_g g(Y)) \in G_f]$$
$$= \frac{c_f}{c_g M}$$

Definition 2.3.5. (Rejection Sampling Algorithm):

Let $f, g \in \mathcal{D}(\mathbb{R}^d)$, $c_f, c_g > 0$ and M > 0. Put $\tilde{f} = c_f f$ and $\tilde{g} = c_g g$. Suppose that $\tilde{f} \leq M\tilde{g}$. We define the **rejection sampling algorithm** as follows:

- (1) sample $Y \sim g$ and $U \sim \mathrm{Uni}(0,1)$ independently
- (2) if $U \leq \frac{\tilde{f}(Y)}{M\tilde{g}(Y)}$, accept Y, else return to (1).

If we sample $(X_n)_{n\in\mathbb{N}}$ independently using the rejection sampler, then the previous exercises imply that $(X_n)_{n\in\mathbb{N}} \stackrel{iid}{\sim} f$ and the acceptance rate is $\frac{c_f}{c_g M}$.

Note 2.3.6. Phrasing the rejection sampler in terms of \tilde{f} and \tilde{g} instead of f and g is usefule because we may not always be able to solve for the normalizing constants.

3. Decision Theory

3.1. Introduction.

Note 3.1.1. We employ the following notation and conventions:

- data space: a measurable space $(\mathcal{X}, \mathcal{F}_{\mathcal{X}})$
- parameter space: a measurable space $(\Theta, \mathcal{F}_{\Theta})$
- distribution familiy: $(P_{\theta})_{\theta \in \Theta} \subset \mathcal{P}(X, \mathcal{F}_{\mathcal{X}})$
- estimation space: a measurable space $(\mathcal{E}, \mathcal{F}_{\mathcal{E}})$

Definition 3.1.2. Let $\eta:\Theta\to\mathcal{E}$. Then η is said to be an **estimand** if η is $(\mathcal{F}_{\Theta},\mathcal{F}_{\mathcal{E}})$ -measurable.

Definition 3.1.3. Let $\eta: \Theta \to \mathcal{E}$ be an estimand and $\delta: \mathcal{X} \to \mathcal{E}$. Then δ is said to be an **estimator of** η if δ is $(\mathcal{F}_{\mathcal{X}}, \mathcal{F}_{\mathcal{E}})$ -measurable. We denote the set of estimators for η by Δ_{η} .

Definition 3.1.4. Let $\eta: \Theta \to \mathcal{E}$ be an estimand and $L: \Theta \times \mathcal{E} \to [0, \infty)$. Then L is said to be a **loss function for** η if

- (1) $L(\theta, \cdot)$ is $(\mathcal{F}_{\mathcal{E}}, \mathcal{B}(\mathbb{R}))$ -measurable
- (2) for each $\theta \in \Theta$, $L(\theta, \eta(\theta)) = 0$

Definition 3.1.5. Let $\eta: \Theta \to \mathcal{E}$ be an estimand and $L: \Theta \times \mathcal{E} \to [0, \infty)$ be a loss function for η . We define the **risk function associated to** L, denoted $R_L: \Theta \times \Delta_{\eta} \to [0, \infty)$, by

$$R_L(\theta, \delta) = \int_{\mathcal{X}} L(\theta, \delta(x)) dP_{\theta}(x)$$

Definition 3.1.6. Let $\eta: \Theta \to \mathcal{E}$ be an estimand, $L: \Theta \times \mathcal{E} \to [0, \infty)$ be a loss function for η and $\Pi \in \mathcal{P}(\Theta, \mathcal{F}_{\Theta})$.

3.2. Bayes Risk.

Definition 3.2.1. Let $\eta: \Theta \to \mathcal{E}$ be an estimand, $L: \Theta \times \mathcal{E} \to [0, \infty)$ be a loss function for η and $\Pi \in \mathcal{P}(\Theta, \mathcal{F}_{\Theta})$. We define the **Bayes risk for** L and Π , denoted $r_{L,\Pi}: \Delta_{\eta} \to [0, \infty)$, by

$$r_{L,\Pi}(\delta) = \int_{\Theta} R_L(\theta, \delta) d\Pi(\theta)$$

Definition 3.2.2. Let $\eta: \Theta \to \mathcal{E}$ be an estimand, $L: \Theta \times \mathcal{E} \to [0, \infty)$ be a loss function for η , $\Pi \in \mathcal{P}(\Theta, \mathcal{F}_{\Theta})$ and $\delta^* \in \Delta_{\eta}$. Then δ^* is said to be a **Bayes estimator for** L and Π if

$$r_{L,\Pi}(\delta^*) = \inf_{\delta \in \Delta_{\eta}} r_{L,\Pi}(\delta)$$

3.3. Minimax Estimation.

Definition 3.3.1. Let $\eta: \Theta \to \mathcal{E}$ be an estimand, $L: \Theta \times \mathcal{E} \to [0, \infty)$ be a loss function for η and $\delta^* \in \Delta_{\eta}$. Then δ^* is said to be a **minimax estimator for** η **and** L if

$$\sup_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}, \boldsymbol{\delta}^*) = \inf_{\boldsymbol{\delta} \in \Delta_{\boldsymbol{\eta}}} \sup_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}, \boldsymbol{\delta})$$

4. Posterior Consistency

4.1. Introduction.

Definition 4.1.1. Let $(\mathcal{X}, \mathcal{F})$ and Θ be