1

Introduction to Statistics

Carson James

1

Contents

Notation

$\mathcal{M}_+(X, \mathcal{A})$	finite measures on $(X, \mathcal{A})$
v	velocity

Preface

0.1 Introduction

Definition 0.1.0.1. Let $A \in \mathcal{B}(R^d)$ and $\Theta \neq \varnothing$. Suppose that $m(A) > 0$. We define

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

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

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

0.2 Sampling

0.2.1 Inverse CDF Sampling

0.2.2 Importance Sampling

0.2.3 Rejection Sampling

Exercise 0.2.3.1. 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

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

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

Exercise 0.2.3.2. 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 0.2.3.3. (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 $U \sim \mathrm{Uni}(0,1)$ are independent, then $(X, cUf(X)) \sim \mathrm{Uni}(G_c)$.
2. If $(X, V) \sim \mathrm{Uni}(G_c)$, then $X \sim f$.

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

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

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

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

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

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

So $X \sim f$.

□

Exercise 0.2.3.4. 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

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

Since $\{(Y, MUc_g g(Y)) \in G_f\} = \{U \leq \frac{c_f f(Y)}{Mc_g g(Y)}\}$, a previous exercise tells us that

$$(Y, MUc_g g(Y))|U \leq \frac{c_f f(Y)}{Mc_g g(Y)} \sim \text{Uni}(G_f)$$

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

$$Y|U \leq \frac{c_f f(Y)}{Mc_g g(Y)} \sim f$$

Finally we have that

$$\begin{aligned} P\left(U \leq \frac{c_f f(Y)}{Mc_g g(Y)}\right) &= P[(Y, MUc_g g(Y)) \in G_f] \\ &= \frac{c_f}{c_g M} \end{aligned}$$

□

Definition 0.2.3.5. (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 $U \sim \text{Uni}(0, 1)$ 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 $(X_n)_{n\in\mathbb{N}} \overset{iid}{\sim} f$ and the acceptance rate is $\frac{c_f}{c_g M}$.

Note 0.2.3.6. 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.

0.3 Decision Theory

0.3.1 Introduction

Note 0.3.1.1. 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: $(P_\theta)_{\theta \in \Theta} \subset \mathcal{P}(\mathcal{X}, \mathcal{F}_\mathcal{X})$
- estimation space: a measurable space $(\mathcal{E}, \mathcal{F}_\mathcal{E})$

Definition 0.3.1.2. 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 0.3.1.3. 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 0.3.1.4. 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 0.3.1.5. 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 0.3.1.6. Let $\eta : \Theta \to \mathcal{E}$ be an estimand, $L : \Theta \times \mathcal{E} \to [0, \infty)$ be a loss function for η and $\Pi \in \mathcal{P}(\Theta, \mathcal{F}_\Theta)$.

0.3.2 Bayes Risk

Definition 0.3.2.1. 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 0.3.2.2. 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

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

0.3.3 Minimax Estimation

Definition 0.3.3.1. 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

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

0.4 Posterior Consistency

0.4.1 Introduction

Definition 0.4.1.1. Let $(\mathcal{X}, \mathcal{F})$ and Θ be

0.5 Frechet Mean

0.5.1 Introduction

Definition 0.5.1.1. Let $(\mathcal{M}, d)$ be a metric space and $\mu \in M_+(\mathcal{M})$. Suppose that $d \in L^2(\mu)$.

- We define the **Frechet function of** μ, denoted $F^d_\mu : \mathcal{M} \to \mathbb{R}$ by

$$F^d_\mu(x) := \int_{\mathcal{M}} d(x, y)^2 \, d\mu(y).$$

- We define the **Fechet mean set of** μ **with respect to** d, denoted $\mathcal{F}(\mu, d) \subset \mathcal{M}$, by $\mathcal{F}(\mu, d) := \underset{x \in \mathcal{M}}{\arg\min}\, F^d_\mu(x)$.

- Let $x_0 \in \mathcal{M}$. Then x_0 is said to be a **Frechet mean of** μ **with respect to** d if $x_0 \in \mathcal{F}(\mu, d)$.

Definition 0.5.1.2. Let $(\mathcal{M}, d)$ be a metric space and $\mu \in M_1(\mathcal{M})$. Suppose that $d \in L^2(\mu)$. Let $(x_n)_{n\in\mathbb{N}} \subset \mathcal{M}$. Let $n \in \mathbb{N}$. Define $\mu_n \in M_1(\mathcal{M})$ by $\mu_n := \frac{1}{n} \sum\limits_{j=1}^{n} \delta_{x_j}$.

- We define the n**-th sample Frechet function of** μ_n, denoted $F^d_{\mu_n} : \mathcal{M} \to \mathbb{R}$ by

$$F^d_{\mu_n}(x) := \int_{\mathcal{M}} d(x, y)^2 \, d\mu_n(y).$$

- Let $x_0 \in \mathcal{M}$. Then x_0 is said to be an n**-th sample Frechet mean of** $(x_j)_{j\in\mathbb{N}}$ **with respect to** d if $x_0 \in \mathcal{F}(\mu_n, d)$.

Note 0.5.1.3. We recall the projection map $\pi : U(n, k) \to U(n, k)/U(k)$ that make and cite exercise in section about steifel and grassmann manifolds implies that $f : U(n, k) \to G(n, k)$ defined by and $f(U) := UU^*$ is a quotient map and $\operatorname{Im} f \cong U(n, k)/U(k)$.

Exercise 0.5.1.4. Let $(\Sigma_j)_{j\in\mathbb{N}} \subset G(n, k)$. Let $N \in \mathbb{N}$. Define $\mu^{(N)} \in M_1(G(n, k))$ by $\mu^{(N)} := \frac{1}{N} \sum\limits_{j=1}^{N} \delta_{\Sigma_j}$. Define $\bar{\Sigma}^{(N)} \in G(n, k)$ by $\bar{\Sigma}^{(N)} := \frac{1}{N} \sum\limits_{j=1}^{N} \Sigma_j$. There exists $\bar{V}^{(N)} \in U(n)$ and $\bar{\Lambda}^{(N)} \in S(n)$ such that $\bar{\Lambda}^{(N)}$ is diagonal and $\bar{\Sigma}^{(N)} = \bar{V}^{(N)} \bar{\Lambda}^{(N)} (\bar{V}^{(N)})^*$. Suppose that $\bar{\Lambda}^{(N)} = \operatorname{diag}(\bar{\lambda}^{(N)}_1, \ldots, \bar{\lambda}^{(N)}_n)$ and $\bar{\lambda}^{(N)}_1 \geq \cdots \geq \bar{\lambda}^{(N)}_1 \geq 0$. Define $U^{(N)} \in U(n, k)$ and $\Sigma^{(N)} \in G(n, k)$ by $U^{(N)} := \bar{V}^{(N)} I_{n,k}$ and $\Sigma^{(N)} := U^{(N)} (U^{(N)})^*$. Then $\Sigma^{(N)} \in \mathcal{F}(\mu^{(N)}, d_{G(n,k)})$.

Proof. We note that for each $\Sigma \in G(n,k)$,

$$\begin{aligned}
F^{d_{G(n,k)}}_{\mu^{(N)}}(\Sigma) &= \int_{\mathcal{M}} d_{G(n,k)}(\Sigma,\Pi)^2 \, d\mu^{(N)}(\Pi) \\
&= \frac{1}{N}\sum_{j=1}^{N} d_{G(n,k)}(\Sigma,\Sigma_j)^2 \\
&= \frac{1}{N}\sum_{j=1}^{N} \|\Sigma - \Sigma_j\|_F^2 \\
&= \frac{1}{N}\sum_{j=1}^{N} \|(\Sigma - \bar{\Sigma}^{(N)}) + (\bar{\Sigma}^{(N)} - \Sigma_j)\|_F^2 \\
&= \frac{1}{N}\sum_{j=1}^{N} \left[\|\Sigma - \bar{\Sigma}^{(N)}\|_F^2 + \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2 + 2\,\mathrm{tr}[(\Sigma - \bar{\Sigma}^{(N)})^*(\bar{\Sigma}^{(N)} - \Sigma_j)]\right] \\
&= \frac{1}{N}\sum_{j=1}^{N} \left[\|\Sigma - \bar{\Sigma}^{(N)}\|_F^2 + \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2 + 2\,\mathrm{tr}[\Sigma^*\bar{\Sigma}^{(N)} - \Sigma^*\Sigma_j + (\bar{\Sigma}^{(N)})^*\bar{\Sigma}^{(N)} - (\bar{\Sigma}^{(N)})^*\Sigma_j]\right] \\
&= \frac{1}{N}\sum_{j=1}^{N} \left[\|\Sigma - \bar{\Sigma}^{(N)}\|_F^2 + \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2\right] + 2\,\mathrm{tr}[\Sigma^*\bar{\Sigma}^{(N)} - \Sigma^*\bar{\Sigma}^{(N)} + (\bar{\Sigma}^{(N)})^*\bar{\Sigma}^{(N)} - (\bar{\Sigma}^{(N)})^*\bar{\Sigma}^{(N)}] \\
&= \frac{1}{N}\sum_{j=1}^{N} \left[\|\Sigma - \bar{\Sigma}^{(N)}\|_F^2 + \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2\right] \\
&= \frac{1}{N}\sum_{j=1}^{N} \|\Sigma - \bar{\Sigma}^{(N)}\|_F^2 + \frac{1}{N}\sum_{j=1}^{N} \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2 \\
&= \|\Sigma - \bar{\Sigma}^{(N)}\|_F^2 + \frac{1}{N}\sum_{j=1}^{N} \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2 \\
&= \left[\|\Sigma\|_F^2 + \|\bar{\Sigma}^{(N)}\|_F^2 + 2\,\mathrm{tr}(\Sigma^*\bar{\Sigma}^{(N)})\right] + \frac{1}{N}\sum_{j=1}^{N} \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2 \\
&= k + \|\bar{\Sigma}^{(N)}\|_F^2 + 2\,\mathrm{tr}(\Sigma^*\bar{\Sigma}^{(N)}) + \frac{1}{N}\sum_{j=1}^{N} \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2.
\end{aligned}$$

Then

$$\begin{aligned}
\operatorname*{arg\,max}_{\Sigma \in G(n,k)} F^{d_{G(n,k)}}_{\mu^{(N)}}(\Sigma) &= \operatorname*{arg\,max}_{\Sigma \in G(n,k)} \left[k + \|\bar{\Sigma}^{(N)}\|_F^2 + 2\,\mathrm{tr}(\Sigma^*\bar{\Sigma}^{(N)}) + \frac{1}{N}\sum_{j=1}^{N} \|\bar{\Sigma}^{(N)} - \Sigma_j\|_F^2\right] \\
&= \operatorname*{arg\,max}_{\Sigma \in G(n,k)} \mathrm{tr}(\Sigma^*\bar{\Sigma}^{(N)}).
\end{aligned}$$

Since

$$\max_{\Sigma \in G(n,k)} \mathrm{tr}(\Sigma^*\bar{\Sigma}^{(N)}) = \max_{U \in U(n,k)} \mathrm{tr}(UU^*\bar{\Sigma}^{(N)}),$$

an exercise in the analysis notes section on optimizing over compact groups implies that

$$\begin{aligned}
\Sigma^{(N)} &\in \operatorname*{arg\,max}_{\Sigma \in G(n,k)} \mathrm{tr}(\Sigma^*\bar{\Sigma}^{(N)}) \\
&= \operatorname*{arg\,max}_{\Sigma \in G(n,k)} F^{d_{G(n,k)}}_{\mu^{(N)}}(\Sigma) \\
&= \mathcal{F}(\mu^{(N)}, d_{G(n,k)}).
\end{aligned}$$

$\square$