Notes

Victor Trappler

Directeurs de Thèse: Arthur VIDARD (Inria)
Élise ARNAUD (UGA)
Laurent DEBREU (Inria)

March 30, 2020

Forward, inverse problems and probability theory

Model space data space and forward problem

We are going to follow Tarantola's description of model and data space in Tarantola [2005].

In order to describe accurately a physical system, we have to define the notion of model:

Definition 1.1 – Model: A model $\mathfrak{M}$ is defined as a pair composed of a forward operator $\mathcal{M}$, and a parameter space Θ

$$\mathfrak{M} = (\mathcal{M}, \Theta)$$

The forward operator is the mathematical representation of the physical system, while the parameter space is chosen here to be a subset of a finite dimensional space, so usually Θ will be a subset of $\mathbb{R}^n$.

Remark 1.2: The dimension of a model $\mathfrak{M} = (\mathcal{M}, \Theta)$ is the number of parameters not reduced to a singleton, so if $\Theta \subset \mathbb{R}^n$, the dimension of $\mathfrak{M}$ is $d \leq n$

Example 1.3: A model with parameter space $\Theta = \mathbb{R}^2 \times [0, 1]$ has dimension 3, while $\Theta = \mathbb{R}^2 \times \{1\}$ has dimension 2.

The data space is formally introduced as the set of all possible observations that one can make during the physical experiment, so consists in all the physically acceptable results of the physical experiment. This set is noted $\mathbb{Y}$. Then, the forward operator $\mathcal{M}$ maps the parameter space $\Theta \subset \mathbb{R}^d$ to the data space $\mathbb{Y}$, as one can expect that all models provide physically acceptable outputs.

Forward problem

Given a model $(\mathcal{M}, \Theta)$, the forward problem consists in applying the forward operator to a given $\theta \in \Theta$, in order to get the model prediction. The forward problem is then to obtain information on the result of the experiment, based on the parameters:

$$\begin{array}{rccc} \mathcal{M}: & \Theta & \longrightarrow & \mathbb{Y} \\ & \theta & \longmapsto & \mathcal{M}(\theta) \end{array}$$

Inverse Problem

The inverse problem, as its name suggests, consists in trying to gather more information on the parameters, based on the result of the experiment, or the physical process.

In that perspective we are going to introduce briefly the usual probabilistic framework.

Probabilistic formulation

Notions of probabilities theory

We are first going to define the usual notions of probability theory, such as events, random variables and density functions. Let us consider a probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$, and a measurable state (or sample) space $(S, \mathcal{B}(S))$.

> **Definition 1.4 – Event, and probability of an event:** We call an event an element of the σ-algebra $\mathcal{F}$, and the probability of an event $A \in \mathcal{F}$ is defined as the Lebesgue integral
>
> $$\mathbb{P}[A] = \int_A \mathrm{d}\mathbb{P}(\omega) \tag{1}$$

Observing an event $B \in \mathcal{F}$ can bring information upon another event $A \in \mathcal{F}$. In that sense, we introduce the conditional probability of A given B:

> **Definition 1.5 – Conditional Probability:** Let A, $B \in \mathcal{F}$ The event A given B

is written $A|B$ and its probability is

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \tag{2}$$

Formally, an event can be seen as an outcome of some uncertain experiment

Definition 1.6 – Random Variable: A random variable (abbreviated as r.v.) Y is a measurable function from $\Omega \longrightarrow S$. A random variable will usually be written with an uppercase letter

Remark 1.7: When $S = \mathbb{R}^p$ with $p > 1$, and $\mathcal{B}(\mathbb{R}^p)$ the usual borelian σ-algebra on $\mathbb{R}^p$, a random variable is called a random vector.

Definition 1.8 – Expectation of a r.v.: The expectation of a r.v. $Y : \Omega \to S$ is defined as

$$\mathbb{E}[Y] = \int_\Omega Y(\omega)\, \mathrm{d}\mathbb{P}(\omega)$$

Remark 1.9: Using the definition 1.8, the probability of an event A can be seen as the expectation of a well chosen random variable:

$$\begin{array}{rccc} \mathbb{1}_A : & \Omega & \longrightarrow & \{0,1\} \\ & \omega & \longmapsto & \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases} \end{array}$$

and

$$\begin{aligned} \mathbb{E}[\mathbb{1}_A] &= \int_\Omega \mathbb{1}_A\, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_A \mathrm{d}\mathbb{P}(\omega) = \mathbb{P}[A] \end{aligned}$$

$\mathbb{1}_A$ is called the indicator function of the event A.

Definition 1.10 – Image (Pushforward) measure: Let $Y : \Omega \to S$ be a random variable, and $A \subseteq S$. The image measure (also called pushforward measure) of $\mathbb{P}$ through Y is denoted by $\mathbb{P}_Y = \mathbb{P} \circ Y^{-1}$. This notation can differ slightly depending on the community in which it is applied, so one can find $\mathbb{P}_Y = \mathbb{P} \circ Y^{-1} = Y_\sharp \mathbb{P}$, the latter notation especially used in transport theory. The probability, for the r.v. Y to

be in A is equal to

$$\mathbb{P}[Y \in A] = \mathbb{P}_Y[A] = \int_A \mathrm{d}\mathbb{P}_Y(\omega) = \int_{Y^{-1}(A)} \mathrm{d}\mathbb{P}(\omega) = \mathbb{P}[Y^{-1}(A)] = \mathbb{P}[\{\omega\,;\, Y(\omega) \in A\}]$$

Real-valued random variables We are now going to focus on real-valued random variables, so measurable function from Ω to the sample space $(S, \mathcal{B}(S)) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 1.11 – Cumulative distribution function (c.d.f.) and Probability density function (p.d.f.): The *cumulative distribution function* (further abbreviated as cdf) of a real-valued r.v. Y is defined as the probability of the right closed intervals that generate the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ of the real line.

$$F_Y(y) = \mathbb{P}\left[Y \leq y\right] = \mathbb{P}_Y\left[\,]-\infty; y]\,\right]$$

and $\lim_{-\infty} F_Y = 0$ and $\lim_{+\infty} F_Y = 1$

If the pushforward measure $\mathbb{P}_Y$ is absolutely continuous with respect to the Lebesgue measure λ defined as $\lambda\left(]a, b]\right) = b - a$, then according to Radon-Nikodym theorem, there exists a function p_Y, such that for all measurable set A,

$$\mathbb{P}_Y[A] = \mathbb{P}[Y \in A] = \int_A \mathrm{d}\mathbb{P}_Y(\omega) = \int_A p_Y(y)\,\mathrm{d}y$$

This function $p_Y : S \subseteq \mathbb{R} \to \mathbb{R}$, is called the *probability density function* (abbreviated pdf) of Y, called the Radon-Nikodym derivative of $\mathbb{P}_Y$ wrt λ: $p_Y = \frac{\mathrm{d}\mathbb{P}_Y}{\mathrm{d}\lambda} = \frac{\mathrm{d}F_Y}{\mathrm{d}y}$. As Y is real-valued, the probability for Y to be in an interval is

$$\mathbb{P}_Y\left[\,]a;\, b]\right] = \mathbb{P}[a \leq Y < b] = \int_a^b p_Y(y)\,\mathrm{d}y = F_Y(b) - F_Y(a)$$

and $\mathbb{P}_Y[\mathbb{R}] = \int_{\mathbb{R}} p_Y(y)\,\mathrm{d}y = 1$. By slight abuse of notation, the notation $\mathbb{P}_Y\,[]$ will be used to indicate that the probability is to be taken wrt Y

Remark 1.12: If the random variable Y admits a density p_Y, its expectation can be rewritten as

$$\mathbb{E}\left[Y\right] = \int_\Omega Y(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}} y\,\mathrm{d}\mathbb{P}_Y(y) = \int_{\mathbb{R}} y\,p_Y(y)\,\mathrm{d}y$$

using the change of variable $Y(\omega) = y$

Definition 1.13 – Quantile function: For a r.v. Y, the quantile function Q_Y is the generalized inverse function of the cdf:

$$Q_y(p) = \inf\{q : F_Y(q) \geq p\}$$

Real-valued random vectors

Definition 1.14 – Joint, marginal and conditional densities: Let $X = [X_1, \cdots, X_p]$ be a random vector from $\Omega \to S \subseteq \mathbb{R}^p$ The cdf of X at the point $x = [x_1, \dots x_p]$ is

$$F_X(x) = F_{X_1,\dots,X_p}(x_1, \dots, x_p) = \mathbb{P}\left[X_1 \leq x_1, \cdots, X_p \leq x_p\right] = \mathbb{P}\left[\bigcap_{i=1}^{p} \{\omega;\, X_i(\omega) \leq x_i\}\right]$$

Similarly as in the real-valued case, we can define the pdf of the random vector, by derivating with respect to the variables:

$$p_X(x) = p_{X_1,\dots,X_p}(x_1, \dots, x_p) = \frac{\partial^p F_X}{\partial x_1 \cdots \partial x_p}(x)$$

and $\int_{\mathbb{R}^p} p_{X_1,\dots,X_p}(x_1, \dots, x_p)\, \mathrm{d}(x_1, \dots, x_p) = 1$

For notation clarity, we are going to set $X = [Y, Z]$ We can now define the marginals

$$p_Y(y) = \int_{\mathbb{R}} p_{Y,Z}(y, z)\, \mathrm{d}z \quad \text{and} \quad p_Z(z) = \int_{\mathbb{R}} p_{Y,Z}(y, z)\, \mathrm{d}y$$

The random variable Y given Z, denoted by $Y \mid Z$ has the conditional density

$$p_{Y|Z}(y \mid z) = \frac{p_{Y,Z}(y, z)}{p_Z(z)}$$

Definition 1.15 – Independence: Let $A, B \in \mathcal{F}$. Those two events are deemed independent if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$. Quite similarly, two real-valued random variables Y and Z are said to be independent if $F_{Y,Z}(y, z) = F_Y(y)F_Z(z)$ or equivalently, $p_{Y,Z}(y, z) = p_Y(y)p_Z(z)$

Bayes' Theorem

Theorem 1.16 – Bayes' theorem:

$$p(\theta \mid y) = \frac{\mathcal{L}(\theta; y)p(\theta)}{p(y)} \propto \mathcal{L}(\theta; y)p(\theta) \tag{3}$$

where

$$p(y) = \int_{\theta \in \Theta} \mathcal{L}(\theta; y)p(\theta)\, \mathrm{d}\theta \tag{4}$$

Measure of the discrepancy

It involves then to compare the output of the mathematical model, with the output of the physical process y, usually by defining a kind of discrepancy D between two elements of $\mathbb{Y}$ verifying:

- $$\begin{array}{rccc} D: & \mathbb{Y} \times \mathbb{Y} & \longrightarrow & \mathbb{R}^+ \\ & (y, y') & \longmapsto & D(y, y') \end{array}$$
- for all y and all y', $D(y, y') \geq 0$
- $D(y, y') = 0 \Rightarrow y = y'$

If $\mathbb{Y}$ is a metric space, one obvious choice for D is to take the associated metric. In the following, unless stated otherwise, $\mathbb{Y}$ will have a finite dimension, and D will be defined as

$$D(y, y') = \|y - y'\|_\Sigma = \sqrt{(y - y')^T \Sigma^{-1} (y - y')} \tag{5}$$

From the physical experiment to the model

The physical system (the reality) that is observed can formally be represented by a model, so by an operator $\mathscr{M}$, applied to a set of parameters $\vartheta \in \Theta_0$:

$$\begin{array}{rccc} \mathscr{M}: & \Theta_0 & \longrightarrow & \mathbb{Y} \\ & \vartheta & \longmapsto & \mathscr{M}(\vartheta) \end{array}$$

Let us assume that we dispose of a model $(\mathcal{M}, \Theta)$, at the input x,

$$\mathscr{M}(x, \vartheta) = \mathcal{M}(x, \theta) + \delta(x, \theta)$$

The difference δ is the error between the physical model and the model. One current assumption, is that $\delta(x, \theta)$ is normally distributed $\delta(\theta) \sim \mathcal{N}(0, \Sigma)$

$$y \mid \theta \sim \mathcal{N}(\mathcal{M}(\theta), \Sigma) \tag{6}$$

Its probability density function can be written as

$$p(y \mid \theta) = (2\pi)^{-n/2} (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2} (\mathcal{M}(\theta) - y)^T \Sigma^{-1} (\mathcal{M}(\theta) - y) \right) \tag{7}$$

Remark 1.17: If $\Sigma = \sigma^2 I$, the pdf can be rewritten as

$$p(y \mid \theta) = (2\pi)^{-n/2}\sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}(\mathcal{M}(\theta) - y)^T(\mathcal{M}(\theta) - y)\right)$$

Definition 1.18: The probability density function of the observation given the parameters is also called the likelihood, $\mathcal{L}$

$$\mathcal{L}(\theta; y) = p(y \mid \theta) \tag{8}$$

Parameter inference

Choosing a likelihood model

$$\begin{aligned} p(y|k, u, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}SS(k, u)\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\|\mathcal{M}(k, u) - y\|_\Sigma^2\right] \end{aligned}$$

Quantities derived from the likelihood

The score function

Given the data y, the likelihood function is $\mathcal{L}(\theta; y)$, and the log likelihood is $l(\theta; y) = \log \mathcal{L}(\theta; y)$ The score function is defined as

$$s(\theta) = \frac{\partial \log \mathcal{L}}{\partial \theta}(\theta; y) \tag{9}$$

and the MLE $\hat{\theta}$ verifies

$$s(\hat{\theta}) = 0 \tag{10}$$

For the true parameter $\bar{\theta}$, averaging over all possible information yields 0:

$$\mathbb{E}\left[s(\bar{\theta}) \mid \bar{\theta}\right] = \mathbb{E}\left[\frac{\frac{\partial p(y|\theta)}{\partial \theta}}{p(y|\theta)} \mid \bar{\theta}\right] = \int \frac{\frac{\partial p(y|\bar{\theta})}{\partial \theta}}{p(y \mid \bar{\theta})} p(y \mid \bar{\theta})\,\mathrm{d}y = \frac{\partial}{\partial \theta}\int p(y|\bar{\theta})\,\mathrm{d}y = 0 \tag{11}$$

The variance of the score is the Fisher Information matrix

Fisher Information Matrix

$$\mathcal{I}(\theta) = \mathbb{E}\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2 \mid \theta\right] \tag{12}$$

$$= \mathbb{E}\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} \mid \theta\right] \tag{13}$$

Priors

Informative priors

$$K \sim \mathcal{U}(\mathbb{K}), \quad p(k)$$
$$U \sim \mathcal{U}(\mathbb{U}), \quad p(u)$$

Non-informative priors

Non-informative priors Now to Bayes' theorem

$$p(k, u|y, \sigma^2) = \frac{p(y|k, u, \sigma^2)p(k, u)}{\iint_{\mathbb{K}\times\mathbb{U}} p(y|k, u, \sigma^2)p(k, u)\,\mathrm{d}(k, u)}$$

Let us assume an hyperprior for σ^2: $p(\sigma^2)$

In the following, we write $\theta = (k, u) \in \Theta$ when no distinction is needed, or a general notation is needed.

Model selection

Likelihood ratio test

The likelihood ratio test is a useful test in the case of nested models, as described in waht follows:

Nested models

Definition 3.1 – Nested models: Let $\mathfrak{M}_1 = (\mathcal{M}_1, \Theta_1)$ and $\mathfrak{M}_2 = (\mathcal{M}_2, \Theta_2)$ be two models. $\mathfrak{M}_1$ is said to be nested within $\mathfrak{M}_2$ if

$$\mathcal{M}_1 = \mathcal{M}_1 \text{ and } \Theta_1 \subset \Theta_2$$

Example 3.2: Let us consider two models, where $\mathbb{Y} = \mathbb{R}$

$$\mathfrak{M}_1 = ((a,b) \mapsto ab; \quad (a,b) \in \mathbb{R} \times [0;2])$$
$$\mathfrak{M}_2 = \big((a,b) \mapsto ab; \quad (a,b) \in \mathbb{R}^+ \times \{1/\pi\}\big)$$

$\mathfrak{M}_2$ is nested within $\mathfrak{M}_1$

Example 3.3: Now let us consider $\mathbb{Y}$ as the space of random vector of dimension n:

$$\mathfrak{M}_1 : (X, A, \sigma) \mapsto AX + \sigma\epsilon, \text{ with } (X, A, \sigma) \in \mathbb{R}^n \times \mathbb{R}^{n\times n} \times \mathbb{R}^+ \text{ and } \epsilon \sim \mathcal{N}(0, I)$$
$$\mathfrak{M}_2 : (X, A, \sigma) \mapsto AX + \sigma\epsilon, \text{ with } (X, A, \sigma) \in \mathbb{R}^n \times \mathbb{R}^{n\times n} \times \{1\} \text{ and } \epsilon \sim \mathcal{N}(0, I)$$

$\mathfrak{M}_2$ is nested within $\mathfrak{M}_2$

Using the likelihood defined above, we can test for the following hypotheses:

- $\mathcal{H}_0$: $\theta \in \Theta_0 \subset \mathbb{R}^d$
- $\mathcal{H}_1$: $\theta \in \Theta_1 \subset \mathbb{R}^r$, and $\Theta_0 \subset \Theta_1$

Intuitively, we can see Θ_1 as the more general model. The test statistic is

$$\Lambda(y) = \frac{\sup_{\theta\in\Theta_0} \mathcal{L}(\theta; y)}{\sup_{\theta\in\Theta_1} \mathcal{L}(\theta; y)} \tag{14}$$

and under $\mathcal{H}_0$, the quantity

$$-2\log\Lambda(y) \xrightarrow{\text{d}} \chi^2_{r-d} \tag{15}$$

is asymptotically distributed as a χ^2_{r-d}. Using the log-likelihood, $-2(l(\theta_0; y) - l(\theta_1; y)) \xrightarrow{\text{d}} \chi^2_{r-d}$ The asymptotic rejection region of level α is then

$$\text{RejReg}_\alpha = \{y \mid -2\log\Lambda(y) > \chi^2_{1-\alpha,r-d}\} \tag{16}$$
$$= \{y \mid \log\Lambda(y) < -\frac{1}{2}\chi^2_{1-\alpha,r-d}\} \tag{17}$$
$$= \{y \mid (\sup_{\theta\in\Theta_0} l(\theta; y) - \sup_{\theta\in\Theta_1} l(\theta; y)) < -\frac{1}{2}\chi^2_{1-\alpha,r-d}\} \tag{18}$$
$$= \{y \mid (\sup_{\theta\in\Theta_1} l(\theta; y) - \sup_{\theta\in\Theta_0} l(\theta; y)) > \frac{1}{2}\chi^2_{1-\alpha,r-d}\} \tag{19}$$
$$\tag{20}$$

Let us set $\theta = (k, u, \phi)$ where ϕ represents additional parameters in the likelihood

$$\mathcal{L}(\theta; y) = \mathcal{L}(k, u, \phi; y) \tag{21}$$

Let us assume furthermore that the maximizer of the likelihood depends only on u (and implicitly on the data).

$$\arg\max_{k\in\mathbb{K}} \mathcal{L}(k,u,\phi) = k^*(u) = \arg\max_{k\in\mathbb{K}} \ell(k,u,\phi) \tag{22}$$

Now let us consider the ratio depending on the uncertain variable u (the y is ignored for notational reasons):

$$-2\log\Lambda(u,\phi') = -2\left(\ell(k,u,\phi) - \ell(k^*(u),u,\phi')\right) \tag{23}$$

Given u, let us define the following likelihoods

$$\mathcal{L}(k;u,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{J(k,u)}{2\sigma^2}\right] \tag{24}$$

$$\mathcal{L}(k=k^*(u);u,\varsigma^2) = \frac{1}{\sqrt{2\pi}\varsigma}\exp\left[-\frac{J^*(u)}{2\varsigma^2}\right] \tag{25}$$

$$\tag{26}$$

Taking the ratio yields

$$\frac{\mathcal{L}(k;u,\sigma^2)}{\mathcal{L}(k^*;u,\varsigma^2)} = \frac{\varsigma}{\sigma}\exp\left[-\frac{1}{2}\left(\frac{J(k,u)}{\sigma^2} - \frac{J^*(u)}{\varsigma^2})\right)\right] \tag{27}$$

$$= \frac{\varsigma}{\sigma}\exp\left[-\frac{1}{2\sigma^2}\left(J(k,u) - \frac{\sigma^2}{\varsigma^2}J^*(u))\right)\right] \tag{28}$$

taking twice the negative log likelihood,

$$-2\log\frac{\mathcal{L}(k;u,\sigma^2)}{\mathcal{L}(k^*;u,\varsigma^2)} = \frac{1}{\sigma^2}\left(J(k,u) - \frac{\sigma^2}{\varsigma^2}J^*(u)\right) + 2\log\frac{\sigma}{\varsigma} \tag{29}$$

The log ratio ϱ is

$$\varrho(k,u,\sigma,\varsigma) = \frac{1}{\sigma^2}\left(J(k,u) - \frac{\sigma^2}{\varsigma^2}J^*(u)\right) + 2\log\frac{\sigma}{\varsigma} \tag{30}$$

$$(\text{When } \sigma=1) = \left(J(k,u) - \frac{1}{\varsigma^2}J^*(u)\right) - 2\log\varsigma \tag{31}$$

Relative Likelihood

Bayesian Model Selection

Let us assume that for $\mathcal{M}$ is chosen to represent the problem at stake. In this case, θ represent implicitly parameters of this model $\mathcal{M}$. Bayes' theorem gives

$$p(\theta|\mathcal{M},y) = \frac{p(y|\mathcal{M},\theta)p(\theta)}{p(y|\mathcal{M})} \tag{32}$$

In Eq.(32), $p(y|\mathcal{M}) = \int_\Theta p(y|\mathcal{M},\theta)p(\theta)\,\mathrm{d}\theta$ is called the evidence of the model $\mathcal{M}$ given the data y.

Bayes factor

When comparing two models $\mathcal{M}_1$ and $\mathcal{M}_2$, one can compute the Bayes factor, that is the ratio of the evidence of the two models:

$$\mathrm{BF}(\mathcal{M}_1, \mathcal{M}_2) = \frac{p(y|\mathcal{M}_1)}{p(y|\mathcal{M}_1)} \tag{33}$$

The generalized normal distribution

Probability density function

We consider a random variable $\xi(\kappa)$ with the following pdf

$$f_\kappa(x, \mu, s) = \frac{\kappa}{2s\Gamma(1/\kappa)} \exp\left[-\left(\frac{|x-\mu|}{s}\right)^\kappa\right] \tag{34}$$

that depends on the parameters μ $s > 0$ and $\kappa > 0$ representing respectively the location, scale and the shape parameter. One can notice that in the particular case where $\kappa = 2$, $\xi(\kappa = 2) \sim \mathcal{N}(\mu, \frac{s^2}{2})$. Similarly, when $\kappa = 1$, $\xi(\kappa = 1)$ is distributed according to Laplace distribution. One important fact is that for $x \in]\mu - s, \mu + s[$, $\frac{|x-\mu|}{s} < 1$ so when $\kappa \to +\infty$, $\exp\left[-\left(\frac{|x-\mu|}{s}\right)^\kappa\right] \to 1$ if $x \in]\mu - s, \mu + s[$, 0 elsewhere. This distribution converges pointwise to a uniform distribution on $]\mu - s, \mu + s[$.

Moments of the generalized normal distribution

Due to the symmetry of the pdf, one can directly conclude that the mode, median and mean are μ:

$$\mathbb{E}[\xi(\kappa)] = \mathrm{Mode}[\xi(\kappa)] = \mathrm{Median}[\xi(\kappa)] = \mu \tag{35}$$

In Pogány and Nadarajah [2010], is also proven the following expression for the variance of $\xi(\kappa)$:

$$\mathbb{V}\mathrm{ar}\left[\xi(\kappa)\right] = s^2\frac{\Gamma(3/\kappa)}{\Gamma(1/\kappa)} = \frac{s^2}{3} - \frac{2s^2\gamma}{3\kappa} + \frac{2s^2(3\gamma^2 + \pi^2)}{9\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right) \tag{36}$$

$$= \frac{(2s)^2}{12} - \frac{2s^2\gamma}{3\kappa} + \frac{2s^2(3\gamma^2 + \pi^2)}{9\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right) \tag{37}$$

At the first order, when $\kappa \to +\infty$, the variance is the variance of a random variable on an interval of length $2s$ as expected

Loglikelihood for GND

$$\ell_\kappa(x, \mu, s) = -\left(\frac{|x-\mu|}{s}\right)^\kappa + \log\kappa - \log(s) - \log\Gamma(1/\kappa) - \log 2 \tag{38}$$

$$= -\left(\left(\frac{x-\mu}{s}\right)^2\right)^{\frac{\kappa}{2}} + \log\kappa - \log(s) - \log\Gamma(1/\kappa) - \log 2 \tag{39}$$

Ratio between two GND

Let us consider two GND distribution: the ratio between the two can be written as

$$\frac{f_{\kappa_1}(x_1,\mu_1,s_1)}{f_{\kappa_2}(x_2,\mu_2,s_2)} = \frac{\kappa_1}{\kappa_2}\frac{s_2\Gamma(1/\kappa_2)}{s_1\Gamma(1/\kappa_1)}\exp\left[\left(\frac{|x_2-\mu_2|}{s_2}\right)^{\kappa_2} - \left(\frac{|x_1-\mu_1|}{s_1}\right)^{\kappa_1}\right] \tag{40}$$

The Profile Likelihood

The likelihood is defined as

$$\mathcal{L}(k,u;y) = p(y \mid k,u) \tag{41}$$

Maximizing the likelihood yields the Maximum Likelihood Estimator: Given the observation y,

$$(k_{\text{MLE}},u_{\text{MLE}}) = \max_{k,u}\mathcal{L}(k,u;y) \tag{42}$$

The traditional profile likelihood is obtained by profiling the nuisance parameters:

$$\mathcal{L}_p(k) = \sup_{u\in\mathbb{U}}\mathcal{L}(k,u;y) \tag{43}$$

Immediately, one can see that maximizing the profile likelihood leads to the MLE.

GP, RR-based family of estimators

Random processes

Let us assume that we have a map f from a p dimensional space to $\mathbb{R}$:

$$\begin{array}{rccc} f: & \mathbb{X} \subset \mathbb{R}^p & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & f(x) \end{array} \tag{44}$$

This function is assumed to have been evaluated on a design of n points, $\mathcal{X} \subset \mathbb{X}^n$. We wish to have a probabilistic modelling of this function We introduce random processes as way to have a prior distribution on function This uncertainty on f is modelled as a random process:

$$\begin{array}{rccc} Z: \mathbb{X} \times \Omega & \longrightarrow & \mathbb{R} \\ (x, \omega) & \longmapsto & Z(x, \omega) \end{array} \tag{45}$$

The ω variable will be omitted next.

Linear Estimation

A linear estimation $\hat{Z}$ of f at an unobserved point $x \notin \mathcal{X}$ can be written as

$$\hat{Z}(x) = \begin{bmatrix} w_1 \dots w_n \end{bmatrix} \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} = \mathbf{W}^T f(\mathcal{X}) = \sum_{i=1}^{n} w_i(x) f(x_i) \tag{46}$$

Using those kriging weights $\mathbf{W}$, a few additional conditions must be added, in order to obtain the Best Linear Unbiased Estimator:

- Non-biased estimation: $\mathbb{E}[\hat{Z}(x) - Z(x)] = 0$
- Minimal variance: $\min \ \mathbb{E}[(\hat{Z}(x) - Z(x))^2]$

Translating using Eq.(46):

$$\mathbb{E}[\hat{Z}(x) - Z(x)] = 0 \iff m(\sum_{i=1}^{n} w_i(x) - 1) = 0 \iff \sum_{i=1}^{n} w_i(x) = 1 \iff \mathbf{1}^T \mathbf{W} = 1 \tag{47}$$

For the minimum of variance, we introduce the augmented vector $\mathbf{Z}_n(x) = [Z(x_1), \dots Z(x_n), Z(x)]$, and the variance can be expressed as:

$$\mathbb{E}[(\hat{Z}(x) - Z(x))^2] = \mathrm{Cov}\left[[\mathbf{W}^T, -1] \cdot \mathbf{Z}_n(x)\right] \tag{48}$$

$$= [\mathbf{W}^T, -1] \, \mathrm{Cov}\left[\mathbf{Z}_n(x)\right] [\mathbf{W}^T, -1]^T \tag{49}$$

In addition, we have

$$\mathrm{Cov}\left[\mathbf{Z}_n(x)\right] = \begin{bmatrix} \mathrm{Cov}\left[\left[Z(x_1) \dots Z(x_n)\right]^T\right] & \mathrm{Cov}\left[\left[Z(x_1) \dots Z(x_n)\right]^T, Z(x)\right] \\ \mathrm{Cov}\left[\left[Z(x_1) \dots Z(x_n)\right]^T, Z(x)\right]^T & \mathbb{V}\mathrm{ar}\left[Z(x)\right] \end{bmatrix} \tag{50}$$

Once expanded, the kriging weights solve then the following optimisation problem:

$$\min_{\mathbf{W}} \mathbf{W}^T \operatorname{Cov}\left[Z(x_1)\dots Z(x_n)\right]\mathbf{W} \tag{51}$$

$$- \operatorname{Cov}\left[\left[Z(x_1)\dots Z(x_n)\right]^T, Z(x)\right]^T \mathbf{W} \tag{52}$$

$$- \mathbf{W}^T \operatorname{Cov}\left[\left[Z(x_1)\dots Z(x_n)\right]^T, Z(x)\right] \tag{53}$$

$$+ \mathbb{V}\mathrm{ar}\left[Z(x)\right] \tag{54}$$

$$\text{s.t.} \mathbf{W}^T\mathbf{1} = \mathbf{1} \tag{55}$$

This leads to

$$\begin{bmatrix}\mathbf{W}\\ m\end{bmatrix} = \begin{bmatrix}\operatorname{Cov}\left[Z(x_1)\dots Z(x_n)\right] & \mathbf{1}\\ \mathbf{1}^T & 0\end{bmatrix}^{-1} \begin{bmatrix}\operatorname{Cov}\left[\left[Z(x_1)\dots Z(x_n)\right]^T, Z(x)\right]^T\\ 1\end{bmatrix} \tag{56}$$

$$= \begin{bmatrix}C(x_1,x_1) & \cdots & C(x_1,x_n) & 1\\ C(x_2,x_1) & \cdots & C(x_2,x_n) & 1\\ \vdots & \ddots & \vdots & \vdots\\ C(x_n,x_1) & \cdots & C(x_n,x_n) & 1\\ 1 & \cdots & 1 & 0\end{bmatrix}^{-1} \begin{bmatrix}C(x_1,x)\\ C(x_2,x)\\ \vdots\\ C(x_n,x)\\ 1\end{bmatrix} \tag{57}$$

Covariance functions

- Desired properties
  - isotropy (?)
  - stationarity
  - semi-definite positiveness
- parametric models of covariance
- examples
- usual hyperparameters estimation

General SUR strategies

Generalities on SUR strategies

Exploration and Space Filling objectives

Contour Estimation

Let ξ be a random process over $\mathbb{X}$, and let us follow what has been done in Bect et al. [2012]. Let ξ_n be the GP constructed using n evaluations of the objective function.

GP of the penalized cost function Δ_α

GP processes

Let $\Delta_\alpha(\mathbf{k},\mathbf{u}) = J(\mathbf{k},\mathbf{u}) - \alpha J^*(\mathbf{u})$. Furthermore, we assume that we constructed a GP on J on the joint space $\mathbb{K}\times\mathbb{U}$, based on a design of n points $\mathcal{X} = \left\{(\mathbf{k}^{(1)},\mathbf{u}^{(1)}),\ldots,(\mathbf{k}^{(n)},\mathbf{u}^{(n)})\right\}$, denoted as $(\mathbf{k},\mathbf{u}) \mapsto Y(\mathbf{k},\mathbf{u})$.

As a GP, Y is described by its mean function m_Y and its covariance function $C(\cdot,\cdot)$, while $\sigma_Y^2(\mathbf{k},\mathbf{u}) = C((\mathbf{k},\mathbf{u}),(\mathbf{k},\mathbf{u}))$

$$Y(\mathbf{k},\mathbf{u}) \sim \mathcal{N}\left(m_Y(\mathbf{k},\mathbf{u}), \sigma_Y^2(\mathbf{k},\mathbf{u})\right) \tag{58}$$

Let us consider now the conditional minimiser:

$$J^*(\mathbf{u}) = J(\mathbf{k}^*(\mathbf{u}),\mathbf{u}) = \min_{\mathbf{k}\in\mathbb{K}} J(\mathbf{k},\mathbf{u}) \tag{59}$$

Analogous to J and J^*, we define Y^* as

$$Y^*(\mathbf{u}) \sim \mathcal{N}\left(m_Y^*(\mathbf{u}), \sigma_Y^{2,*}(\mathbf{u})\right) \tag{60}$$

where

$$m_Y^*(\mathbf{u}) = \min_{\mathbf{k}\in\mathbb{K}} m_Y(\mathbf{k},\mathbf{u}) \tag{61}$$

The surrogate conditional minimiser is used in Ginsbourger profiles etc. The α-relaxed difference Δ_α modelled as a GP can then be written as

Considering the joint distribution of $Y(\mathbf{k},\mathbf{u})$ and $Y^*(\mathbf{u}) = Y(\mathbf{k}^*(\mathbf{u}),\mathbf{u})$, we have

$$\begin{bmatrix} Y(\mathbf{k},\mathbf{u}) \\ Y^*(\mathbf{u}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_Y(\mathbf{k},\mathbf{u}) \\ m_Y^*(\mathbf{u}) \end{bmatrix}; \begin{bmatrix} C\left((\mathbf{k},\mathbf{u}),(\mathbf{k},\mathbf{u})\right) & C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) \\ C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) & C\left((\mathbf{k}^*(\mathbf{u}),\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) \end{bmatrix}\right) \tag{62}$$

By multiplying by the matrix $\begin{bmatrix} 1 & -\alpha \end{bmatrix}$ yields

$$\Delta_\alpha(\mathbf{k},\mathbf{u}) \sim \mathcal{N}\left(m_\Delta(\mathbf{k},\mathbf{u}); \sigma_\Delta^2(\mathbf{k},\mathbf{u})\right) \tag{63}$$

$$m_\Delta(\mathbf{k},\mathbf{u}) = m_Y(\mathbf{k},\mathbf{u}) - \alpha m_Y^*(\mathbf{u}) \tag{64}$$

$$\sigma_\Delta^2(\mathbf{k},\mathbf{u}) = \sigma_Y^2(\mathbf{k},\mathbf{u}) + \alpha^2\sigma_{Y^*}^2(\mathbf{k},\mathbf{u}) - 2\alpha C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) \tag{65}$$

Assuming that $C((\mathbf{k},\mathbf{u}),(\mathbf{k}',\mathbf{u}')) = s\prod_{i\in\mathcal{I}_\mathbf{k}} \rho_{\theta_i}(\|k_i - k_i'\|)\prod_{j\in\mathcal{I}_\mathbf{u}} \rho_{\theta_j}(\|u_j - u_j'\|)$

$$C\left((\mathbf{k},\mathbf{u}),(\mathbf{k}^*(\mathbf{u}),\mathbf{u})\right) = s\prod_{i\in\mathcal{I}_\mathbf{k}} \rho_{\theta_i}(\|k_i - k_i^*(\mathbf{u})\|)\prod_{j\in\mathcal{I}_\mathbf{u}} \rho_{\theta_j}(0) \tag{66}$$

$$= s\prod_{i\in\mathcal{I}_\mathbf{k}} \rho_{\theta_i}(\|k_i - k_i^*(\mathbf{u})\|) \tag{67}$$

Approximation of the objective probability using GP

We are going now to use a different notation for the probabilities, taken with respect to the GP: $\mathcal{P}$, to represent the uncertainty encompassed by the GP.

Defined somewhere else, we have

$$\Gamma_\alpha(\mathbf{k}) = \mathbb{P}_{\mathbf{U}}\left[J(\mathbf{k},\mathbf{U}) \leq \alpha J^*(\mathbf{U})\right] \tag{68}$$

$$= \mathbb{E}_{\mathbf{U}}\left[\mathbb{1}_{J(\mathbf{k},\mathbf{U})\leq\alpha J^*(\mathbf{U})}\right] \tag{69}$$

This classification problem can be approached with a plug-in approach, or a probablistic one:

$$\mathbb{1}_{J(\mathbf{k},\mathbf{u})\leq\alpha J^*(\mathbf{u})} \approx \mathbb{1}_{m_Y(\mathbf{k},\mathbf{u})\leq\alpha m_Y^*(\mathbf{u})} \tag{70}$$

$$\mathbb{1}_{J(\mathbf{k},\mathbf{u})\leq\alpha J^*(\mathbf{u})} \approx \mathcal{P}\left[\Delta_\alpha(\mathbf{k},\mathbf{u}) \leq 0\right] = \pi_\alpha(\mathbf{k},\mathbf{u}) \tag{71}$$

Using the GPs, for a given $\mathbf{k}$, α and $\mathbf{u}$, the probability for our meta model to verify the inequality is given by Based on those two approximation, the approximated probability Γ is

$$\hat{\Gamma}_{\alpha,n}(\mathbf{k}) = \mathbb{P}_U\left[m_Y(\mathbf{k},\mathbf{u}) \leq \alpha m_Y^*(\mathbf{u})\right] \qquad \text{(plug-in)}$$

$$\hat{\Gamma}_{\alpha,n}(\mathbf{k}) = \mathbb{E}_U\left[\mathcal{P}\left[\Delta_\alpha(\mathbf{k},\mathbf{u}) \leq 0\right]\right] = \mathbb{E}_U\left[\pi_\alpha(\mathbf{k},\mathbf{u})\right] \qquad \text{(Probabilistic approx)} \tag{72}$$

The probability of coverage for the set $\{Y - \alpha Y^*\}$ is π_α, and can be computed using the CDF of the standard normal distribution Φ:

$$\pi_\alpha(\mathbf{k},\mathbf{u}) = \Phi\left(-\frac{m_{\Delta_\alpha}(\mathbf{k},\mathbf{u})}{\sigma_{\Delta_\alpha}(\mathbf{k},\mathbf{u})}\right) \tag{73}$$

Finally, averaging over $\mathbf{u}$ yields

$$\hat{\Gamma}(\mathbf{k}) = \int_{\mathbb{U}} \pi_\alpha(\mathbf{k},\mathbf{u}) p(\mathbf{u})\,\mathrm{d}\mathbf{u} \tag{74}$$

Sources, quantification of uncertainties, and SUR strategy ?

Formally, for a given point $(\mathbf{k},\mathbf{u})$, the event "the point is α-acceptable" has probability $\pi(\mathbf{k},\mathbf{u})$ and variance $\pi(\mathbf{k},\mathbf{u})(1-\pi(\mathbf{k},\mathbf{u}))$. Obviously, the points with the highest uncertainty have the highest variance, so have a coverage probability π around 0.5.

Random sets

Let us start by introducing diverse tools based around Vorob'ev expectation of closed sets (ref thèse Reda), Heinrich et al. [2012].

Let us consider A, a random closed set, such that its realizations are subsets of $\mathbb{X}$, and p is its coverage probability, that is

$$p(x) = \mathbb{P}\left[x \in A\right], x \in \mathbb{X} \tag{75}$$

For $\eta \in [0,1]$, we define the η-level set of p,

$$Q_\eta = \{x \in \mathbb{X} \mid p(x) \geq \eta\} \tag{76}$$

It may seem trivial, but let us still note that those sets are decreasing:

$$0 \leq \eta \leq \xi \leq 1 \implies Q_\xi \subseteq Q_\eta \tag{77}$$

Using those level sets for the level $\eta = 0.05$ for instance:

$$Q_{1-\frac{\eta}{2}} \subset Q_{\frac{\eta}{2}} \tag{78}$$

Recalling the objective, it gives upper bounds and lower bounds of the confidence interval of level η on the probability for each $\mathbf{k}$:

$$\hat{\Gamma_\alpha}^U(\mathbf{k}) = \mathbb{P}_\mathbf{U}\left[x = (\mathbf{k}, \mathbf{u}) \in Q_{1-\frac{\eta}{2}}\right] \tag{79}$$

$$\hat{\Gamma_\alpha}^L(\mathbf{k}) = \mathbb{P}_\mathbf{U}\left[x = (\mathbf{k}, \mathbf{u}) \in Q_{\frac{\eta}{2}}\right] \tag{80}$$

In Dubourg et al. [2011] is introduced the Margin of uncertainty, defined as the following set difference

$$\mathbb{M}_\eta = Q_{\frac{\eta}{2}} \setminus Q_{1-\frac{\eta}{2}} \tag{81}$$

Considering the

Let μ be a Borel σ-finite measure on $\mathbb{X}$. We define Vorob'ev expectation, as the η^*-level set of A verifying

$$\forall \beta < \eta^* \quad \mu(Q_\beta) \leq \mathbb{E}[\mu(A)] \leq \mu(Q_{\eta^*}) \tag{82}$$

that is the level set of p, that has the volume of the mean of the volume of the random set A.

SUR Strategies

The main idea behind Stepwise Uncertainty Reduction is to define a criterion, say κ_n, that encapsulates the epistemic uncertainty, and to minimize this criterion, in order to select the next point:

$$x^{n+1} = \arg\max_{x \in \mathbb{X}} \kappa_n(x) \tag{83}$$

where κ_n depends on $Y \mid \mathcal{X}_n$ This approach is suitable for step by step evaluations.

Integrated Mean square criterion

Sacks et al. [1989] Let us consider that we have a kriging model over $\mathbb{X}$ based on a experimental design $\mathcal{X}$, that is denoted $Y \mid \mathcal{X}$

We define the Integrated Mean Square Error (IMSE) as

$$\text{IMSE}(Y \mid \mathcal{X}) = \int_\mathbb{X} \sigma^2_{Y|\mathcal{X}}(x)\,\mathrm{d}x \tag{84}$$

where

$$Y \mid \mathcal{X} \sim \mathcal{N}(m_{Y|\mathcal{X}}(x), \sigma^2_{Y|\mathcal{X}}(x)) \tag{85}$$

$$x^{n+1} = \underset{x \in \mathbb{X}}{\arg\min}\ \mathbb{E}_{y \sim Y(x)} \left[\text{IMSE} \left(Y \mid \mathcal{X} \cup \{(x, y)\} \right) \right] \tag{86}$$

So we choose th point minimizing the expected integrated mean square error.

Weighted IMSE

To include a more precise objective than the enrichment of the design, one can add a weight function to the integral, giving the $W - \text{IMSE}$:

$$w - \text{IMSE}(Y \mid \mathcal{X}) = \int_{\mathbb{X}} \sigma^2_{Y|\mathcal{X}}(x) w(x) \, \mathrm{d}x \tag{87}$$

In order to increase the accuracy of the surrogate model around some region of interest, the $w - \text{IMSE}$ can be transformed into

$$w - \text{IMSE}(Y \mid \mathcal{X}) = \int_{\mathbb{X}} \sigma^2_{Y|\mathcal{X}}(x) \mathcal{P} \left[x \in \mathbb{M}_\eta \right] \, \mathrm{d}x \tag{88}$$

where $\mathcal{M}_\eta$ is the η-margin of uncertainty.

UB-LB for $(p, \alpha_p, \mathbf{k}_p)$

Let us assume that we have set a probability $p \in [0, 1]$. Let us recall that the triplet $(p, \alpha_p, \mathbf{k}_p)$ verifies

$$\max_{\mathbf{k}} \Gamma_{\alpha_p}(\mathbf{k}) = \Gamma_{\alpha_p}(\mathbf{k}_p) = \mathbb{P}_{\mathbf{U}} \left[J(\mathbf{k}_p, \mathbf{U}) \leq \alpha_p J^*(\mathbf{U}) \mid \mathbf{U} = \mathbf{u} \right] = p \tag{89}$$

Let us say that $\bar{\Gamma}$ is the η-upper-bound, while $\underline{\Gamma}$ is the η-lower bounds, so

$$\mathcal{P} \left[\underline{\Gamma}(\mathbf{k}) \leq \Gamma_n(\mathbf{k}) \leq \bar{\Gamma}(\mathbf{k}) \right] = \eta \tag{90}$$

- If $\underline{\Gamma}(\mathbf{k}) > p$, we are too permissive, so we should decrease α
  - by how much ?
- If $\bar{\Gamma}(\mathbf{k}) < p$, we are too conservative, so we should increase α
  - by how much again ?
- If $\underline{\Gamma}(\mathbf{k}) < p < \bar{\Gamma}(\mathbf{k})$, reduce uncertainty on $\mathbf{k}_p$

Changing the value of α does not require any further evaluation of the objective function, so can be increased until $\max \hat{\Gamma} = p$? by dichotomy for instance. This $\hat{\mathbf{k}}_p$ is then the candidate.

Criterion: stepwise reduction of the variance of the estimation of $\hat{\Gamma}(\hat{\mathbf{k}}_p) = \max_{\mathbf{k}} \hat{\Gamma}(\hat{\mathbf{k}})$

For a fixed $p \in (0, 1]$, and an initial design $\mathcal{X}$. Set an initial value for $\alpha \geq 1$.

- Define Δ_α, using $Y \mid \mathcal{X}$
- Update α such that $\max \hat{\Gamma}_{\alpha,n} = p$
- Compute measure of uncertainty that we want to reduce:
  - $\bar{\Gamma}_{\alpha,n}(\mathbf{k}) - \underline{\Gamma}_{\alpha,n}(\mathbf{k})$
  - $\pi_\alpha(\mathbf{k}, \mathbf{u})(1 - \pi_\alpha(\mathbf{k}, \mathbf{u}))$

References

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, May 2012. ISSN 0960-3174, 1573-1375. doi: 10.1007/s11222-011-9241-4.

V. Dubourg, B. Sudret, and J.-M. Bourinet. Reliability-based design optimization using kriging surrogates and subset simulation. *Structural and Multidisciplinary Optimization*, 44(5):673–690, Nov. 2011. ISSN 1615-147X, 1615-1488. doi: 10.1007/s00158-011-0653-8.

P. Heinrich, R. S. Stoica, and V. C. Tran. Level sets estimation and Vorob'ev expectation of random compact sets. *Spatial Statistics*, 2:47–61, Dec. 2012. ISSN 22116753. doi: 10.1016/j.spasta.2012.10.001.

T. K. Pogány and S. Nadarajah. On the characteristic function of the generalized normal distribution. *Comptes Rendus Mathematique*, 348(3-4):203–206, Feb. 2010. ISSN 1631073X. doi: 10.1016/j.crma.2009.12.010.

J. Sacks, S. B. Schiller, and W. J. Welch. Designs for Computer Experiments. *Technometrics*, 31(1):41–47, Feb. 1989. ISSN 0040-1706, 1537-2723. doi: 10.1080/00401706.1989.10488474.

A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, Pa, 2005. ISBN 978-0-89871-572-9. OCLC: 265659758.