§3 STATISTICAL INVERSION THEORY

so far: forward operator $K : \mathcal{X} \to \mathcal{Y}$ is linked to noisy observation y^δ by equation of form

$$y^\delta = Kx + \eta = y + \eta$$

goal: given $y_{OBS} = y^\delta$, find x or functions thereof

inverse problem: problem of retrieving unknown information of unknown quantities by indirect observations

statistical inference: problem of inferring properties of an unknown distribution from data generated from that distribution

(i) frequentist approach:

- η is modeled as a random variable
  since $y_{OBS} = y^\delta$ is a function of η, y_{OBS} is a random variable
- x is assumed to be deterministic
- an estimate $\tilde{x}$ of x is a function of $y_{OBS} = y^\delta$, and consequently is random

we study properties of the probability distribution of $\tilde{x}$

(ii) Bayesian approach:

- η and the information about x are assumed to be random with some prescribed distributions
  since $y_{OBS} = y^\delta$ is a function of η, y_{OBS} is a random variable
- solution of inverse problem is a distribution

an event is deterministic if its outcome is completely predictable

a random event is complement of a deterministic event
i.e., its outcome is not fully predictable

"randomness" means "lack of information"

example: deterministic event: today is Monday
tomorrow is Tuesday
random event: whether it rains tomorrow is
not fully predictable

Ω is a nonempty set, called the sample space
Ω is equipped with a probability measure $\mathbb{P}$ of events $\mathcal{E} \subset \Omega$
defined on $\mathcal{F}$
$\mathcal{F}$ defines a σ-algebra
sets in $\mathcal{F}$ are called measurable states
a measureable space $(\Omega, \mathcal{F})$ is a set Ω together with
its σ-algebra $\mathcal{F}$
we define $(\Omega, \mathcal{F}, \mathbb{P})$ as a probability space

the probability measure $\mathbb{P}$ on $\mathcal{F}$ is a function on $\mathcal{F}$
with the following properties

$$0 \leq \mathbb{P}[\mathcal{E}] \leq 1, \quad \mathbb{P}[\varnothing] = 0 \quad \mathbb{P}[\Omega] = 1$$

example: fair coin toss; we have $\Omega = \{\text{heads}, \text{tails}\}$,
$\mathcal{F} = \{\varnothing, \{\text{heads}\}, \{\text{tails}\}, \Omega\}$, and $\mathbb{P}[\varnothing] = 0$, $\mathbb{P}[\Omega] = 1$,
$\mathbb{P}[\{\text{heads}\}] = \mathbb{P}[\{\text{tails}\}] = \frac{1}{2}$

probability is additive: if $\mathcal{E}_1, \mathcal{E}_2 \subset \Omega$, $\mathcal{E}_1 \cap \mathcal{E}_2 = \varnothing$,

$$\mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2] = \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2]$$

two events independent if $\mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2] = \mathbb{P}[\mathcal{E}_1]\mathbb{P}[\mathcal{E}_2]$

the conditional probability of ϵ_1 on/given ϵ_2 is probability that ϵ_1 happens provided that ϵ_2 happens

$$\mathbb{P}[\epsilon_1|\epsilon_2] = \frac{\mathbb{P}[\epsilon_1 \cap \epsilon_2]}{\mathbb{P}[\epsilon_2]} \leq 1$$

example:

ϵ_1 event that a bit string of length four contains at least two consecutive 0s

ϵ_2 event that the first bit of a string of length four is a 0

$\mathbb{P}[\epsilon_1|\epsilon_2]$ probability that a bit string of length four has at least two consecutive 0s, given that its first bit is a 0

Because $\epsilon_1 \cap \epsilon_2 = \{0000, 0001, 0010, 0011, 0100\}$, we see that $\mathbb{P}[\epsilon_1 \cap \epsilon_2] = 5/16$

because there are eight bit strings of length four that start with a 0, we have $\mathbb{P}[\epsilon_2] = 8/16 = 1/2$

consequently,

$$\mathbb{P}[\epsilon_1|\epsilon_2] = \frac{5/16}{1/2} = 5/8$$

by definition of independent events we have

$$\mathbb{P}[\epsilon_1|\epsilon_2] = \mathbb{P}[\epsilon_1] \quad \text{and} \quad \mathbb{P}[\epsilon_2|\epsilon_1] = \mathbb{P}[\epsilon_2]$$

if ϵ_1 and ϵ_2 are mutually independent

we will mostly consider probability densities over state space – the set of all possible outcomes

for purpose of this class, we will limit the exposition to the standard Euclidean space $\mathbb{R}^n$

variables that depend the outcomes of random events are called random variables; in the framework of probability they are modeled as functions that are measureable, i.e., a measureable function on a probability space

a random variable X is a mapping from the sample space Ω to the state space S

$$X : \Omega \ni \omega \mapsto X(\omega) \in S$$

more rigorously, a random variable X, taking values in $\mathbb{R}^n$, is a function on $(\Omega, \mathcal{F}, \mathbb{P})$ such that for any measureable set M

$$X^{-1}(M) := \{\omega \in \Omega : X(\omega) \in M\} \in \mathcal{F}$$

in accordance with definition of a measureable function

example: fair die toss ; $S = \{1,2,3,4,5,6\}$

in this case we may define sample space equal to state space, i.e., $\Omega = S$

a particular die is modelled as mapping

$$X : \Omega \to S, \quad i \mapsto i, \quad 1 \le i \le 6$$

and the probability of random events for this die are

$$\mathbb{P}[X(i) = i] = w_i \quad \text{with } \sum_{i=1}^{6} w_i = 1$$

we call $X(\omega)$ a random variable, because we are uncertain about its outcome

we call $x = X(\omega), \omega \in \Omega$, a realization of X

example: suppose that a coin is flipped three times; let $X(\omega)$ be the random variable that equals the number of heads that appear when ω is the outcome; then $X(\omega)$ takes on the following values

$$X(HHH) = 3$$
$$X(THH) = X(HTH) = X(THH) = 2$$
$$X(THT) = X(HTT) = X(TTH) = 1$$
$$X(TTT) = 0$$

example: $X \sim \text{Bernoulli}(p)$

experiment: role a fair die

outcomes: w_i, where $i \in \{1,2,3,4,5,6\}$

sample space: $\Omega = \{1,2,3,4,5,6\}$

random variable:

$$X : \Omega \to S, \qquad X(w_i) = \begin{cases} 1 & \text{if } i \text{ is odd} \\ 0 & \text{if } i \text{ is even} \end{cases}$$

since the outcomes of w_i are random, X takes on the values 0 and 1, randomly;

for the σ-algebra on Ω we use its powerset

$$\mathcal{F} = 2^{\Omega}$$

$$\mathbb{P}[\{w_i\}] = 1/6 \quad \text{for all } i \in \{1,2,3,4,5,6\}$$

$$\mathbb{P}[X=0] = \mathbb{P}[\{w_2, w_4, w_6\}] = 1/2$$

$$\mathbb{P}[X=1] = 1/2$$

distribution $\quad \mathbb{P}_X[M] = 1/2\, \mathbb{I}_X(0) + 1/2\, \mathbb{I}_X(1)$

the distribution of X is the probability measure $\mathbb{P}_X$ on $\mathcal{F}$ which provides the probability that X takes values in any measurable set M

$$\mu_X(M) = \mathbb{P}_X[M] = \mathbb{P} \circ X^{-1}(M) = \mathbb{P}[X(\omega) \in M] \qquad (*)$$

$\mu_X(M)$ is a probability measure

key property of $\mu_X(M)$: probability of event M in state space S to happen ($\mu_X(M) = \mathbb{P}_X[M]$) is equal to probability of event $N = X^{-1}(M)$ in sample space Ω

the probability distribution $\mu_X(M)$ measures the size of the subspace N of Ω mapped onto M by the random variable X

$\mu_X(M)$ can be studied using its cumulative distribution function (CDF) Π_X or its probability density function (PDF) π_X

CDF for random variable $X \in \mathbb{R}$:

$$\Pi_X(x) = \mathbb{P}[X \leq x] \quad \text{for any } x \in \mathbb{R}$$

CDF for random variable $X \in \mathbb{R}^n$:

$$\Pi_X(x) = \mathbb{P}[X \preccurlyeq x] \quad \text{for any } x \in \mathbb{R}^n$$

relation between CDF and PDF for $X \in \mathbb{R}$:

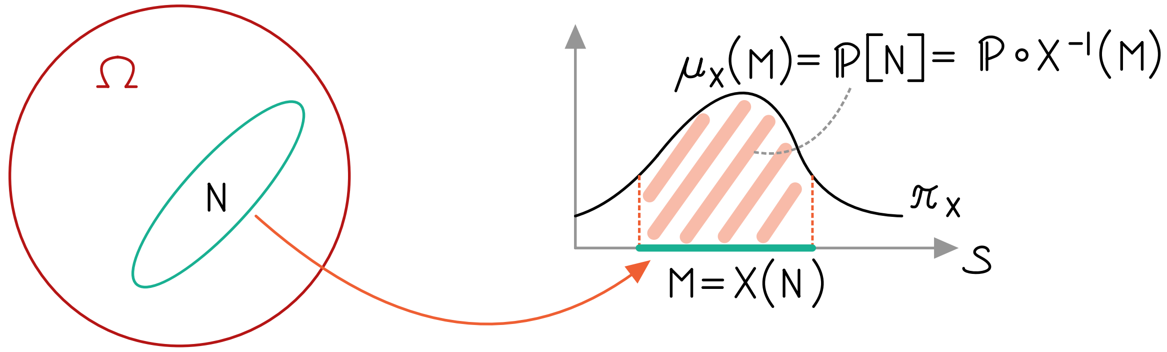
$$\Pi_X(x) = \int_{-\infty}^{x} \pi_X(y)\, dy \quad \text{with} \quad \int_{\mathbb{R}} \pi_X(y)\, dy = 1$$

$\Pi_X(x)$ can be viewed as area under PDF over interval $(-\infty, x]$

the probability distribution $\mu_X(M)$ is linked to the PDF π_X by

$$\mu_X(M) = \int_M \pi_X(x)\, dx \qquad (*)$$

illustration



meaning of random variable: maps events $N \in \Omega$ into set $M = X(N)$ in the state space S such that are under density function $\pi_X(x)$ and above $M = X(N)$ is exactly the probability that N happens

the concepts of probability distribution and probability density introduced in (*) and (*) for random variable X that maps from Ω to S, are valid for (measureable) maps from one arbitrary space to another

the pair $(S, \pi_X(x))$ contains complete information we need for computation; moving forward, we will directly work in state space

remark

for simplicity, we drop dependence on random variable X when possible

$$\pi(x) := \pi_X(x)$$

given a random variable X, the expectation / mean is given by the center of mass of probability distribution

$$\bar{x} = \mathbb{E}[X] = \int_S x\,\pi(x)\,dx = \int_\Omega X(\omega)\,d\omega$$

example: $\Omega = [-1, 1]$; probability $\mathbb{P}$ is uniform

$$\mathbb{P}[\epsilon] = \int_\epsilon dx = \tfrac{1}{2}|\epsilon|, \qquad \epsilon \subset [-1, 1]$$

random variables

$$X_1 : [-1, 1] \to \mathbb{R}, \quad X_1(\omega) = 1 \quad \forall \omega \in \mathbb{R}$$

$$X_2 : [-1, 1] \to \mathbb{R}, \quad X_2(\omega) = \begin{cases} 2, & \omega \geq 0 \\ 0, & \omega < 0 \end{cases}$$

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1$$

the variance is given by the expectation of the squared deviation from the expectation/mean

$$\mathbb{V}\mathrm{ar}[X] = \mathbb{E}[(X-\bar{x})^2] = \int_S (x-\bar{x})^2 \pi(x)\, dx = \int_\Omega (X(\omega) - \bar{x})^2\, d\omega$$

expectation and variance are first two moments of density $\pi(x)$; kth moment:

$$\mathbb{E}[(X-\bar{x})^k] = \int_S (x-\bar{x})^k \pi(x)\, dx$$

third moment: skewness fourth moment: kurtosis

the standard deviation of X is defined as

$$\mathrm{Std}[X] = (\mathbb{V}\mathrm{ar}[X])^{1/2}$$

the covariance of two random variables X and Y with means $\bar{x}$ and $\bar{y}$ is defined as

$$\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\bar{x})(Y-\bar{y})] = \mathbb{E}[XY] - \bar{x}\bar{y}$$

and $\mathrm{Cov}[X,X] = \mathbb{V}\mathrm{ar}[X]$ "mixed central moment"

---

random vectors

the expectation of a random vector $X \in \mathbb{R}^n$ is vector of expectation entry

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}$$

generalizes to matrices

the variance of a random vector $X \in \mathbb{R}^n$ is defined as an $n \times n$ symmetric matrix of all covariances

$$\mathbb{V}\mathrm{ar}[X] = \Sigma \in \mathbb{R}^{n,n}, \quad \Sigma_{ij} = \mathrm{Cov}[X_i, X_j]$$

equivalent definition

$$\mathbb{V}\mathrm{ar}[X] = \mathbb{E}[(X-\bar{x})(X-\bar{x})^T] = \mathbb{E}[XX^T] - \bar{x}\bar{x}^T \in \mathbb{R}^{n,n}$$

with $\bar{x} = \mathbb{E}[X]$

<u>covariance</u> or <u>cross-covariance</u> of $X \in \mathbb{R}^n$ with $Y \in \mathbb{R}^m$

$$\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\bar{x})(Y-\bar{y})^T] = \mathbb{E}[XY^T] - \bar{x}\bar{y}^T \in \mathbb{R}^{n,m}$$

and $\mathrm{Cov}[X,X] = \mathbb{V}\mathrm{ar}[X]$

---

the joint probability distribution of two random variables X and Y with values in S and T defined on the same probability space and joint probability density $\pi_{XY}(x,y) = \pi(x,y)$

$$\begin{aligned}\mu_{XY}(X\in M, Y\in N) &= \mathbb{P}[X\in M, Y\in N] \\ &= \mathbb{P}\circ(X^{-1}(M)\cap Y^{-1}(N)) \\ &= \int_M\int_N \pi_{XY}(x,y)\,dy\,dx\end{aligned}$$

for all $M\times N \subset S\times T$

X and Y are independent if

$$\mu_{XY}(X\in M, Y\in N) = \mu_X(M)\mu_Y(N) \qquad \forall\, M\times N\subset S\times T$$

or if

$$\pi_{XY}(x,y) = \pi_X(x)\,\pi_Y(y)$$

the <u>marginal density</u> of X is the probability of X when Y may take on any value

$$\pi_X(x) = \int_T \pi_{XY}(x,y)\,dy$$

likewise

$$\pi_Y(y) = \int_S \pi_{XY}(x,y)\,dx$$

conditional distributions and conditional expectations play a key role in probability theory and are a basic tool in Bayesian inference from before

the <u>conditional probability</u> of ϵ_1 given ϵ_2 is defined as

$$\mathbb{P}[\epsilon_1|\epsilon_2] = \frac{\mathbb{P}[\epsilon_1 \cap \epsilon_2]}{\mathbb{P}[\epsilon_2]}$$

Bayes' theorem: provides a way to obtain conditional probability of $\mathbb{P}[\epsilon_2|\epsilon_1]$ in terms of $\mathbb{P}[\epsilon_1|\epsilon_2]$

$$\mathbb{P}[\epsilon_2|\epsilon_1] = \frac{\mathbb{P}[\epsilon_1|\epsilon_2]\mathbb{P}[\epsilon_2]}{\mathbb{P}[\epsilon_1]}$$

let X, Y denote discrete random variables taking values $\{x_i\}$ and $\{y_i\}$

$$\pi_{XY}(x_i, y_i) = \mathbb{P}[X=x_i, Y=y_i]$$
$$\pi_Y(y_i) = \mathbb{P}[Y=y_i]$$
$$\pi_{X|Y}(x_i|y_i) = \mathbb{P}[X=x_i|Y=y_i]$$

using Bayes' theorem, we find conditional density of X given Y in terms of that of Y given X

$$\mathbb{P}[X=x_i|Y=y_i] = \frac{\mathbb{P}[X=x_i, Y=y_i]}{\mathbb{P}[Y=y_i]} = \frac{\mathbb{P}[Y=y_i|X=x_i]\,\mathbb{P}[X=x_i]}{\mathbb{P}[Y=y_i]}$$

which can be written in terms of conditional and marginal densities

$$\pi_{X|Y}(x_i|y_i) = \frac{\pi_{XY}(x_i, y_i)}{\pi_Y(y_i)} = \frac{\pi_{X|Y}(x_i|y_i)\,\pi_X(x_i)}{\pi_Y(y_i)}$$

let X, Y denote continuous random variables taking values x and y with joint density π_{XY} and marginal densities π_X and π_Y

the conditional density of X given Y

$$\pi_{X|Y}(x|y) = \frac{\pi_{XY}(x,y)}{\pi_Y(y)} \qquad (*)$$

provided $\pi_Y(y) \neq 0$

X is independent of Y iff

$$\pi_{X|Y}(x|y) = \pi_X(x) \quad \text{or} \quad \pi_{Y|X}(y|x) = \pi_Y(y) \quad \forall x, y$$

follows immediately from $\pi_{XY}(x,y) = \pi_X(x)\,\pi_Y(y)$

Bayes' formula

$$\pi_{X|Y}(x|y) = \frac{\pi_{XY}(x,y)}{\pi_Y(y)} = \frac{\pi_{X|Y}(x|y)\,\pi_X(x)}{\pi_Y(y)} \qquad (*)$$

derivation of (*)

let $C_\delta(y)$ denote a ball of radius δ centered at y; then

$$\mathbb{P}[X \in M, Y \in C_\delta(y)] = \frac{\mathbb{P}[X \in M, Y \in C_\delta(y)]}{\mathbb{P}[Y \in C_\delta(y)]}$$

$$= \frac{\int_{C_\delta(y)} \int_M \pi_{XY}(x,z)\,dx\,dz}{\int_{C_\delta(y)} \pi_Y(z)\,dz}$$

$$= \frac{\frac{1}{|C_\delta(y)|} \int_{C_\delta(y)} \int_M \pi_{XY}(x,z)\,dx\,dz}{\frac{1}{|C_\delta(y)|} \int_{C_\delta(y)} \pi_Y(z)\,dz}$$

shrinking ball to the point y and applying the Lebesgue differentiation theorem yields

$$\lim_{\delta\to 0} \mathbb{P}[X\in M, Y\in C_\delta(y)] = \lim_{\delta\to 0} \frac{\frac{1}{|C_\delta(y)|}\int_{C_\delta(y)}\int_M \pi_{XY}(x,z)\,dx\,dz}{\frac{1}{|C_\delta(y)|}\int_{C_\delta(y)} \pi_Y(z)\,dz}$$

$$= \frac{\int \mathbb{I}_M(x)\,\pi_{XY}(x,y)\,dx}{\pi_Y(y)}$$

which can be shown to lead to

$$\mathbb{P}[X\in M, Y=y] = \frac{\int_M \pi_{XY}(x,y)\,dx}{\pi_Y(y)} = \int_M \pi_{X|Y}(x|y)\,dx$$

which proofs (*) and leads to (*)

the <u>conditional probability density</u> of X given Y is the probability density of X assuming Y takes value y

conditional density $\pi_{X|Y}(x|y)$ leads naturally to conditional expectation

$$\mathbb{E}[X|y] = \int_S x\,\pi_{X|Y}(x|y)\,dx$$