EECE 5550
Mobile Robotics

Lecture 6: Review of Probability

Derya Aksaray

Assistant Professor

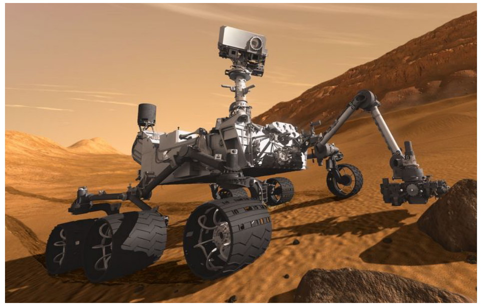
Department of Electrical and Computer Engineering

Northeastern University

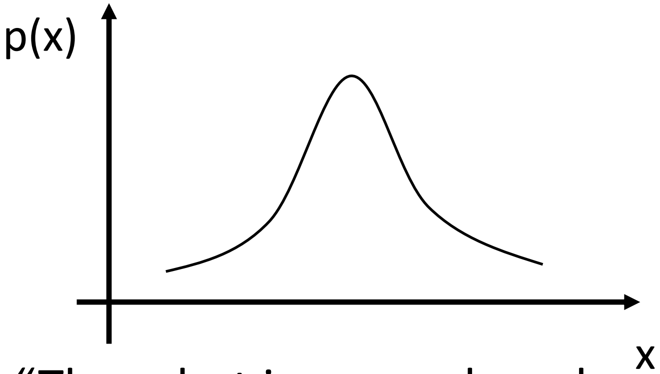
Why probability in robotics?

Most of the time, the state of the robot and state of its environment are unknown and only noisy sensors available.

- Probability provides a framework to fuse sensory information.
- **Result:** Probability distribution over possible states of robot and environment



"The robot is exactly here"

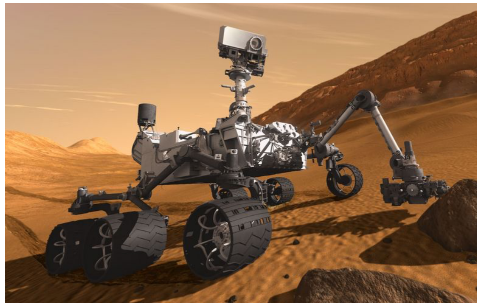


"The robot is somewhere here"

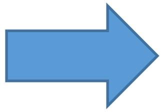
Why probability in robotics?

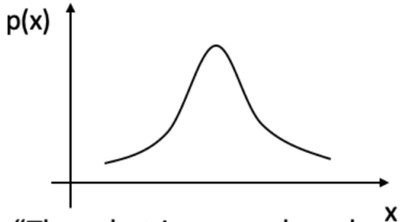
Most of the time, dynamics is often stochastic, hence we can not optimize for a particular outcome, but only optimize to obtain a good distribution over outcomes.

- Probability provides a framework to reason in this setting
- **Result:** Ability to find good control policies for stochastic dynamics and environments



"After taking action i, the robot will exactly be here"





"The robot is somewhere here"

"After taking action i, the robot will be somewhere here"

*One can choose an action that causes zero (or low) probability of collision.

Example 1: Drone

State: position, orientation, velocity, angular rate

Sensors:

- GPS : noisy estimate of position (sometimes also velocity)
- Inertial sensing unit: noisy measurements from
  - 3-axis gyro [=angular rate sensor],
  - 3-axis accelerometer [=measures acceleration + gravity],
  - 3-axis magnetometer

Dynamics:

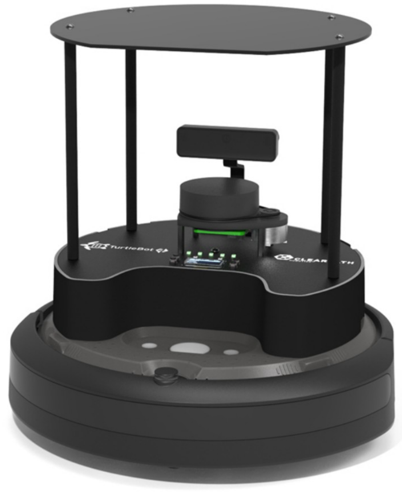
- *Noise:* wind, unmodeled dynamics in engine, servos, blades

Example 2: Mobile ground robot

State: position, heading

Sensors:

- Odometry : sensing motion of actuators (e.g., wheel encoders)
- Laser range finder: sensing distances to obstacles



Dynamics:

- *Noise:* wheel slippage, unmodeled variation on the floor

Basic definitions

A probability space is a triple (Ω, *F*, *P*) consisting of:

- A *sample space* Ω
- An *event space* $F = \{E_k \subseteq \Omega \mid k \epsilon K\}$ whose elements E_k (*events*) are *subsets* of Ω
- A *probability measure* P: $F \rightarrow [0,1]$ that assigns a *probability* $P(E) \epsilon [0,1]$ to each event *E* in *F*.

Example:

Sample space – $\Omega = \{1, 2, 3, 4, 5, 6\}$

Event space - $F = \{\{5\}, \{2,4,6\}\}$

Probability measure - $P(\{5\}) = \frac{1}{6}$, $\quad P(\{2,4,6\}) = \frac{3}{6}$

Basic definitions

A probability space is a triple (Ω, *F*, *P*) consisting of:

- A *sample space* Ω
- An *event space* $F = \{E_k \subseteq \Omega \mid k \in K\}$ whose elements E_k (*events*) are *subsets* of Ω
- A *probability measure* P: $F \rightarrow [0,1]$ that assigns a *probability* $P(E) \in [0,1]$ to each event *E* in *F*.

The measure *P* must satisfy the following two conditions:

- $P(\emptyset) = 0$ and $P(\Omega) = 1$
- **Subadditivity:** If $\{E_k\}$ is any countable set of *disjoint* events (i.e. $E_k \cap E_l = \emptyset$ *for all k ≠ l*), then:

$$P\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} P(E_k)$$

Why think about probability this way?

It provides a common way of thinking about ***both*** discrete ***and*** continuous distributions

Discrete Case

- Sample space Ω is a countable set of *points*
- We measure the probability of *event* $E \subseteq \Omega$ by summing the probabilities of points it contains:

$$P(E) = \sum_{\omega \in E} P(\omega)$$

- Defining object: *probability mass function*

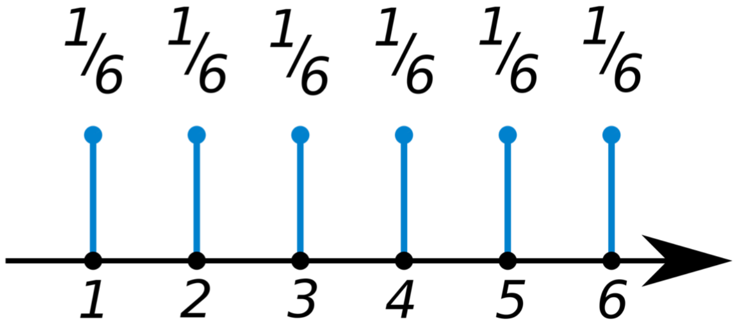
Continuous Case

- Sample space Ω is an uncountable set
- We (typically) measure the probability of an *event* $E \subseteq \Omega$ by integrating a density $p\colon \Omega \to \mathbb{R}$ over *E*:

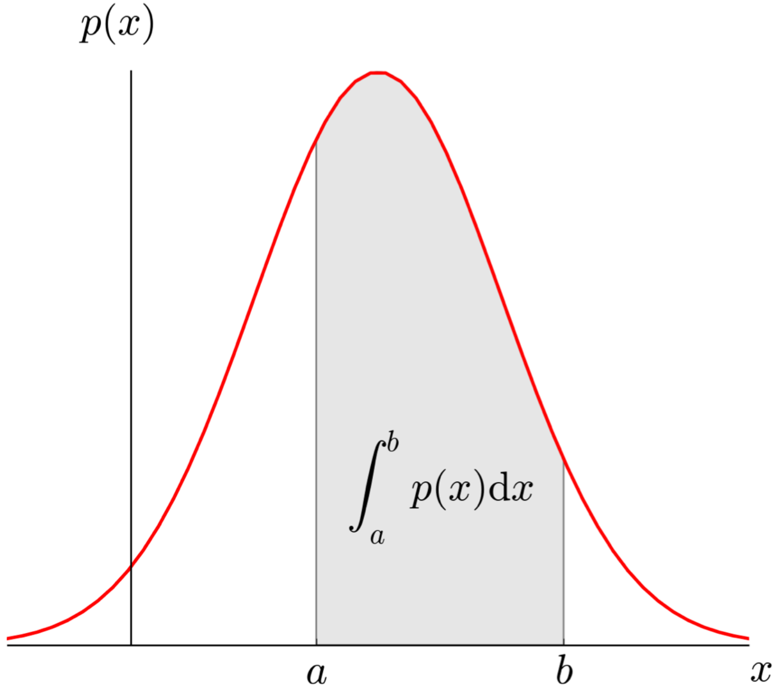
$$P(E) = \int_E p(\omega)\ d\omega$$

- Defining object: *probability density function*

Probability mass/density functions



The probability mass function
of a fair die.



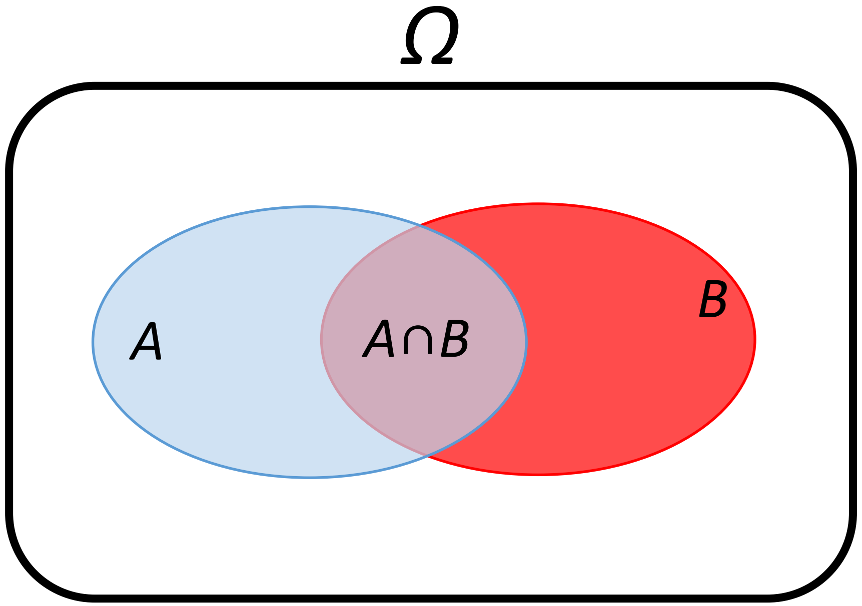
The probability density function
of room temperature.

What is the probability of x=70?

Why think about probability this way?

It implies that statements about "probability" are really just statements about *sizes of sets*

⇒ "Proof by Venn diagram"

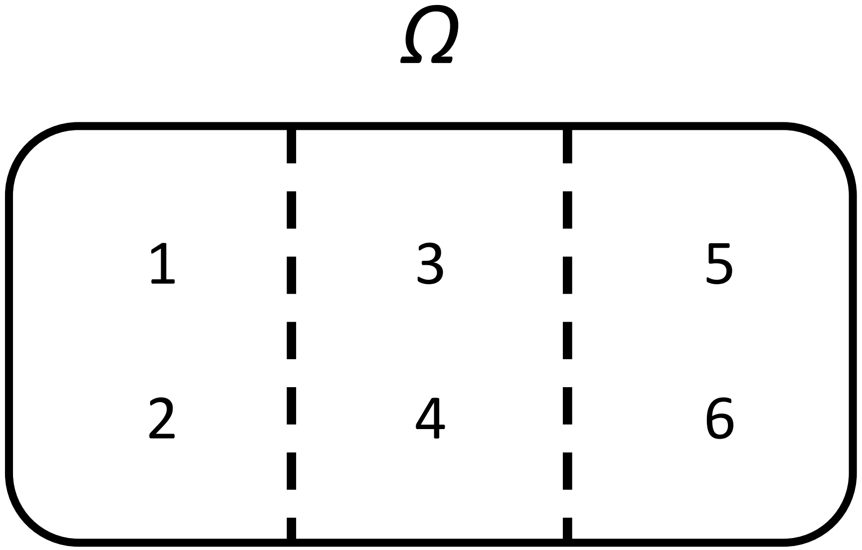


Examples / useful identities:

- **Complement rule:** $P(A) + P(A^c) = 1$
- **Sum rule:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Union bound:** $P(A \cup B) \leq P(A) + P(B)$
- **Law of total probability:** If $\{B_K\}$ is a *countable partition* of Ω, then:

$$P(A) = \sum_k P(A \cap B_k)$$

Example – Law of total probability



Consider three partitions:

- $B_{12} = \{1,2\}$
- $B_{34} = \{3,4\}$
- $B_{56} = \{5,6\}$

Let A be seeing an even number. What is $P(A)$?

$$P(A) = P(A \cap B_{12}) + P(A \cap B_{34}) + P(A \cap B_{56})$$

$$= P(2) + P(4) + P(6)$$

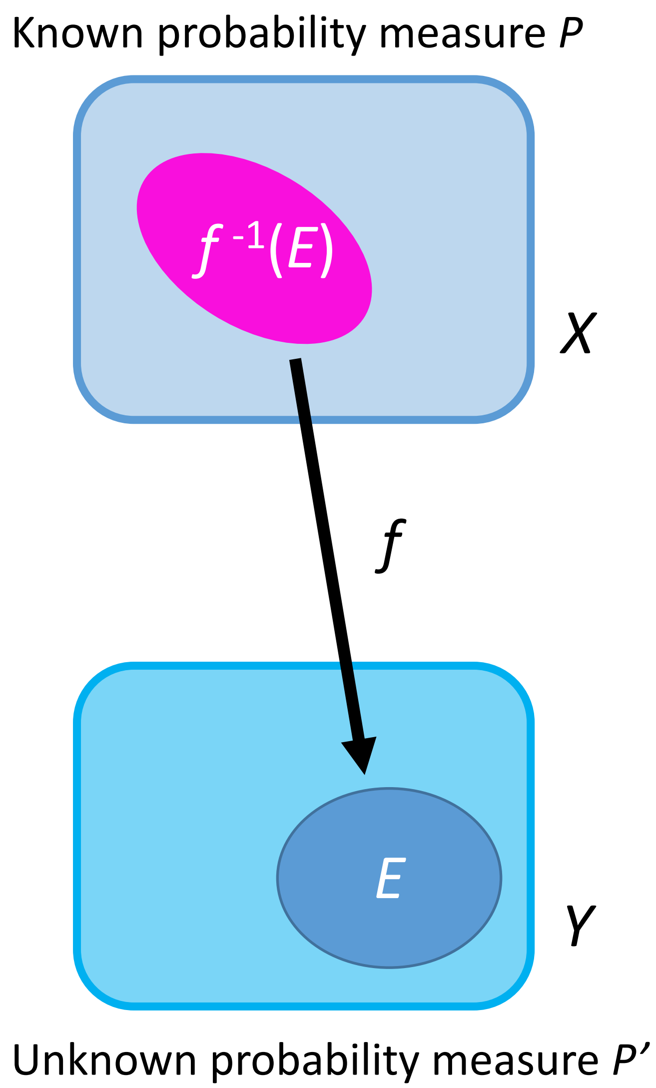
Pushforward measures

Given a sample space X with probability measure P and a function $f: X \to Y$, we can define a new probability measure $f_* P$ on Y according to:

$$f_* P(E) = P(f^{-1}(E))$$

We call $f_* P$ the *pushforward measure* of P by f.

$$f^{-1}(E) = \{x \in X \mid f(x) \in E\}.$$



Pushforward measures

Example:

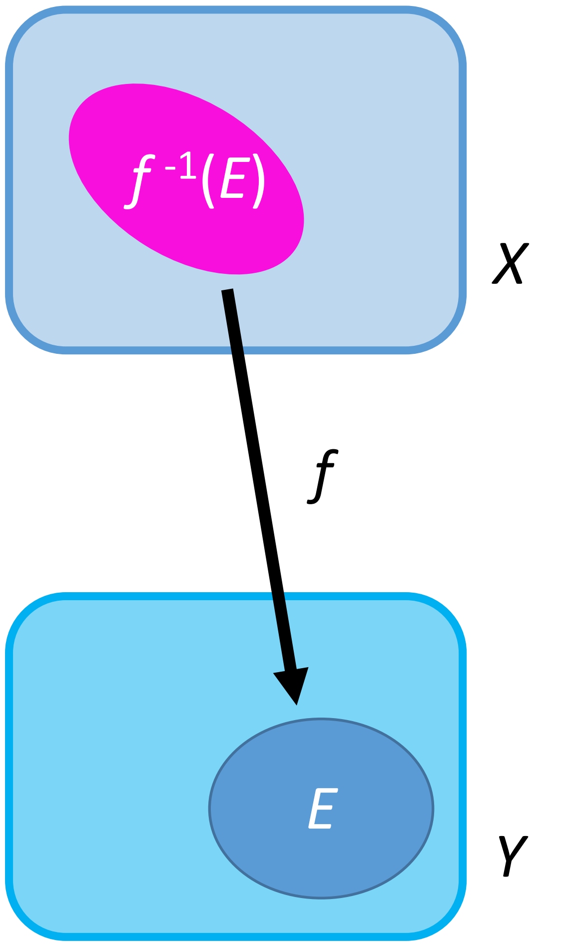
$$X = \{1,2,3,4,5,6\}$$

$$f(1) = f(3) = f(5) = odd$$

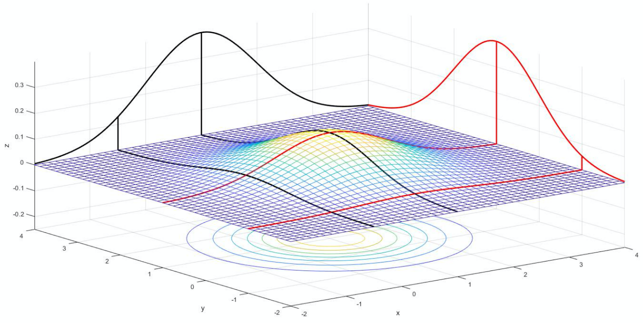
$$f(2) = f(4) = f(6) = even$$

$$Y = \{odd, even\}$$

$$f_*P(even) = P\left(f^{-1}(even)\right) = P(\{2,4,6\})$$

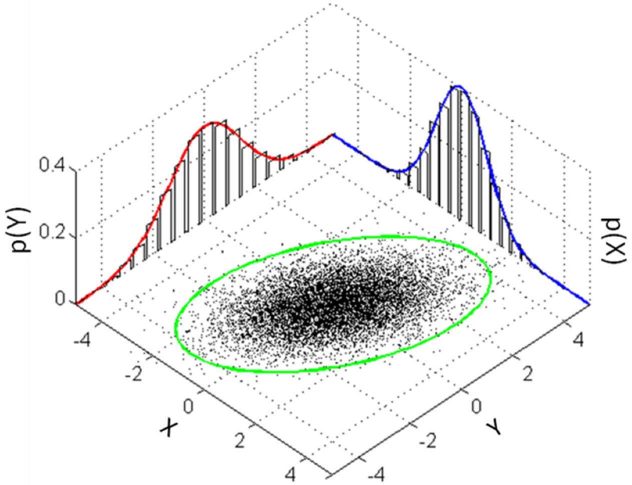


Joint and marginal distributions



Informally:

- **Joint probability:** the probability of two events occurring simultaneously
- **Marginal probability:** the probability of an event irrespective of the outcome of another variable



Let X and Y be two sets, P_{XY} a probability measure on the *product* space $X \times Y$, and $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ the *projection maps* onto X and Y. The pushforward measures:

$$P_X \triangleq (\pi_X)_* P_{XY} \quad \text{and} \quad P_Y \triangleq (\pi_Y)_* P_{XY}$$

on X and Y are called the *marginal* distributions of P_{XY}.

Joint and marginal distributions

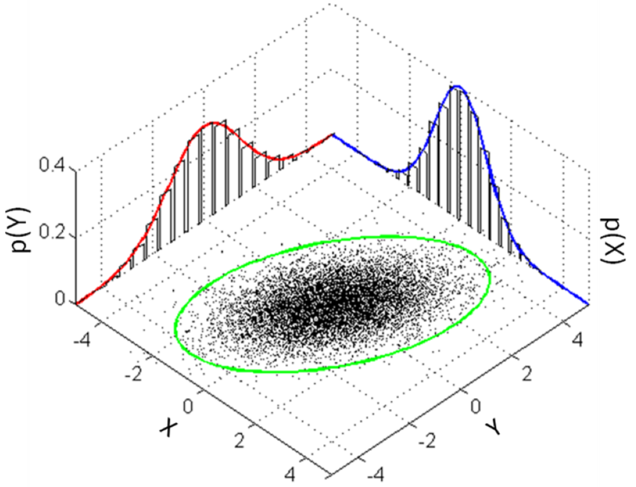
Let X and Y be two sets, P_{XY} a probability measure on the *product* space $X{\times}Y$, and $\pi_X\colon X{\times}Y \to X$ and $\pi_Y\colon X{\times}Y \to Y$ the *projection maps* onto X and Y. The pushforward measures:

$$P_X \triangleq (\pi_X)_* P_{XY} \quad \text{and} \quad P_Y \triangleq (\pi_Y)_* P_{XY}$$

on X and Y are called the *marginal* distributions of P_{XY}.

Special case: If P_{XY} is a continuous probability measure on $\mathbb{R}^m \times \mathbb{R}^n$ with density p_{XY}, then the density p_X of the marginal distribution P_X is given by:

$$p_X(x) = \int_{\mathbb{R}^n} p_{XY}(x, y)\,dy$$

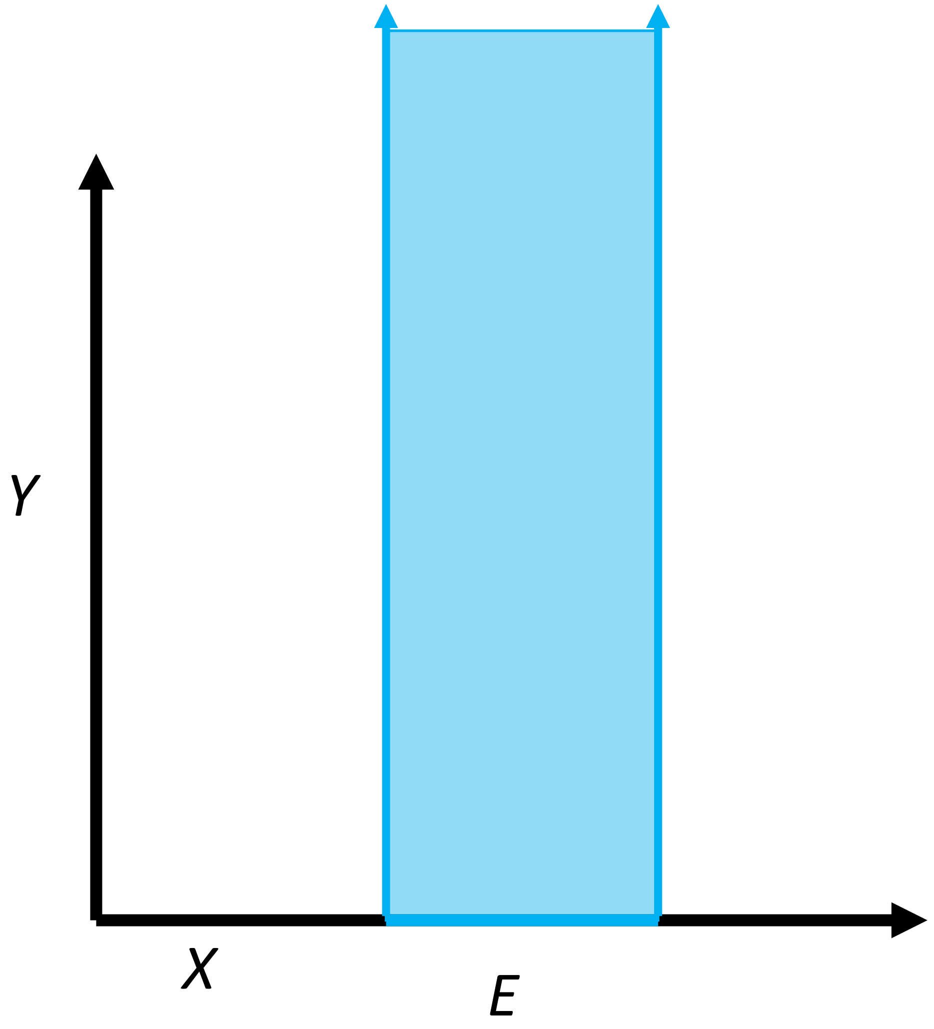


Joint and marginal distributions

Derivation of marginal density:

$$
\begin{aligned}
P_X(E) &\triangleq (\pi_X)_* P_{XY}(E) \\
&= P_{XY}(\pi_X^{-1}(E)) \\
&= P_{XY}(E \times \mathbb{R}^n) \\
&= \int_{E\times\mathbb{R}^n} p_{XY}(x,y)\, dx\, dy \\
&= \int_E \left[\int_{\mathbb{R}^n} p_{XY}(x,y)\; dy \right] dx
\end{aligned}
$$

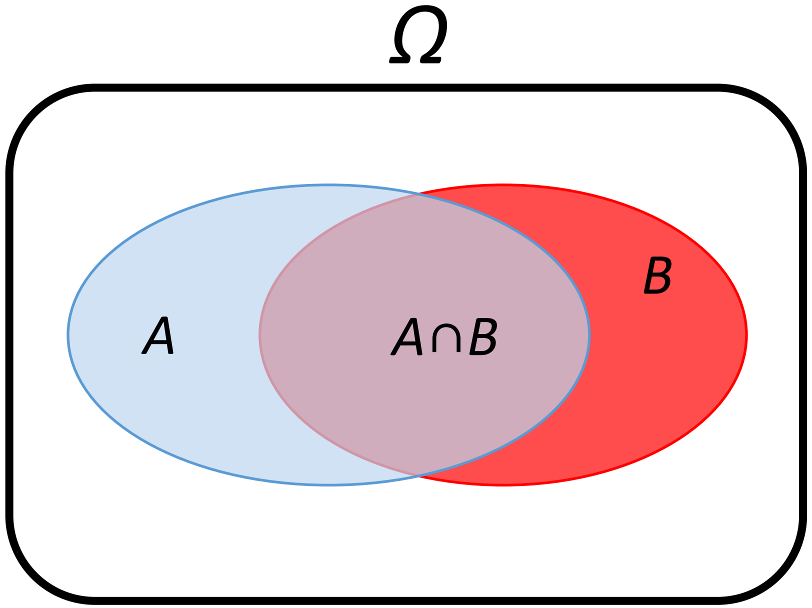
$$\Rightarrow \; p_X(x) = \int_{\mathbb{R}^n} p_{XY}(x,y)\, dy$$



Conditional probability and independence

Given a probability space (Ω, *F*, *P*) and two events *A*,*B* ϵ *F* with *P*(*B*) > 0, the *conditional probability of A given B*, denoted *P*(*A* | *B*), is:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



Intuitively: *P*(*A* | *B*) represents the probability of *A* occurring, given that we know *B* has *already happened*.

Conditional probability and independence

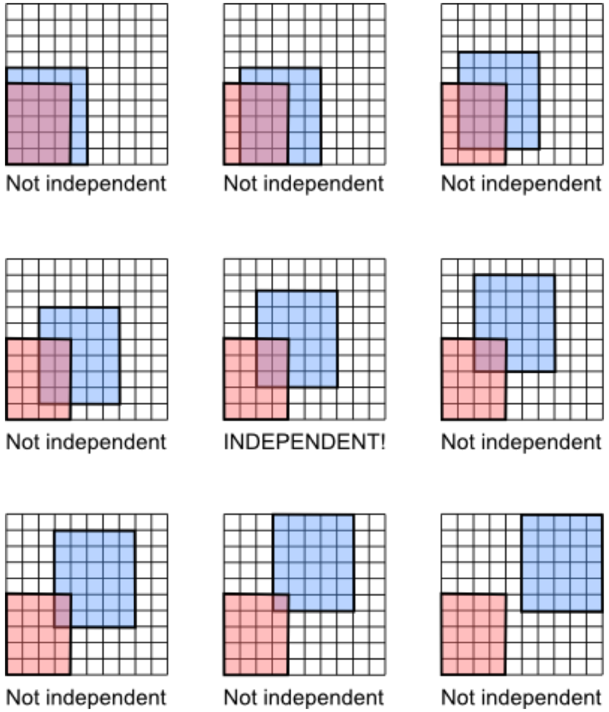
If $P(A \mid B) = P(A)$, then the occurrence of B has no impact on the occurrence of A. In this case, we say that A and B are *independent* events; furthermore, $P(A \cap B) = P(A)P(B)$.

- **Example:**

Head

Tail

the occurrence of one event does not affect the chances of the occurrence of the other event



$A = red$

$B = blue$

$\Pr(A) = ?$

$\Pr(A|B) = ?$

Conditional probability density

Special case: If P_{XY} is a continuous probability measure on $\mathbb{R}^m \times \mathbb{R}^n$ with density p_{XY}, then the density $p_{X|Y}$ of the conditional distribution $P_{X|Y}$ is given by:

$$p_{X|Y}(x \mid Y \in B) = \frac{1}{p(Y \in B)} \int_B p_{XY}(x, y)\, dy$$

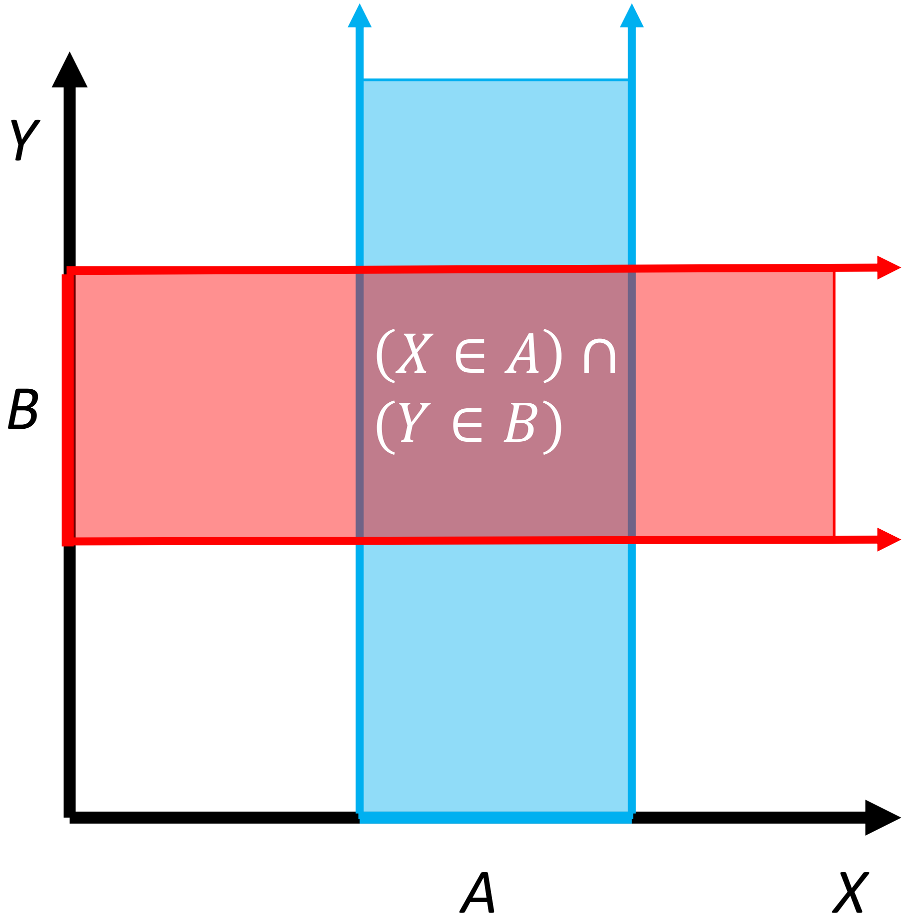
Conditional probability densities

$$P(X \in A \mid Y \in B) = \frac{P\big((X \in A) \cap (Y \in B)\big)}{P(Y \in B)}$$

$$= \frac{1}{P(Y \in B)} \int_{A \times B} p_{XY}(x, y)\, dx\, dy$$

$$= \int_A \left[\frac{1}{P(Y \in B)} \int_B p_{XY}(x, y)\, dy \right] dx$$

$$\Rightarrow \quad p_{X|Y}(x \mid Y \in B) = \frac{1}{P(Y \in B)} \int_B p_{XY}(x, y)\, dy$$



Special case: Given a particular *realization Y = y*, by taking the limit as $B \to \{y\}$ in the above formula, we obtain:

$$p_{X|Y}(x \mid Y = y) = \frac{p_{XY}(x, y)}{p_Y(y)}$$

Example – conditional probability

Let X and Y be two jointly continuous random variables with joint PDF

$$p_{XY}(x,y) = \begin{cases} \frac{x^2}{4} + \frac{y^2}{4} + \frac{xy}{6}, & 0 \le x \le 1, \quad 0 \le y \le 2 \\ 0, & otherwise \end{cases}$$

For $0 \le y \le 2$, find

- The conditional PDF of X given $Y = y$;
- $P(X < \frac{1}{2} \mid Y = y)$

Example - solution

- The marginal PDF of Y for $0 \leq y \leq 2$:

$$p_Y(y) = \int_0^1 \frac{x^2}{4} + \frac{y^2}{4} + \frac{xy}{6} dx = \frac{3y^2 + y + 1}{12}$$

Then,

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)} = \frac{3x^2 + 3y^2 + 2xy}{3y^2 + y + 1}$$

Overall,

$$p_{X|Y}(x|y) = \begin{cases} \dfrac{3x^2 + 3y^2 + 2xy}{3y^2 + y + 1}, & 0 \leq x \leq 1 \\ 0, & otherwise \end{cases}$$

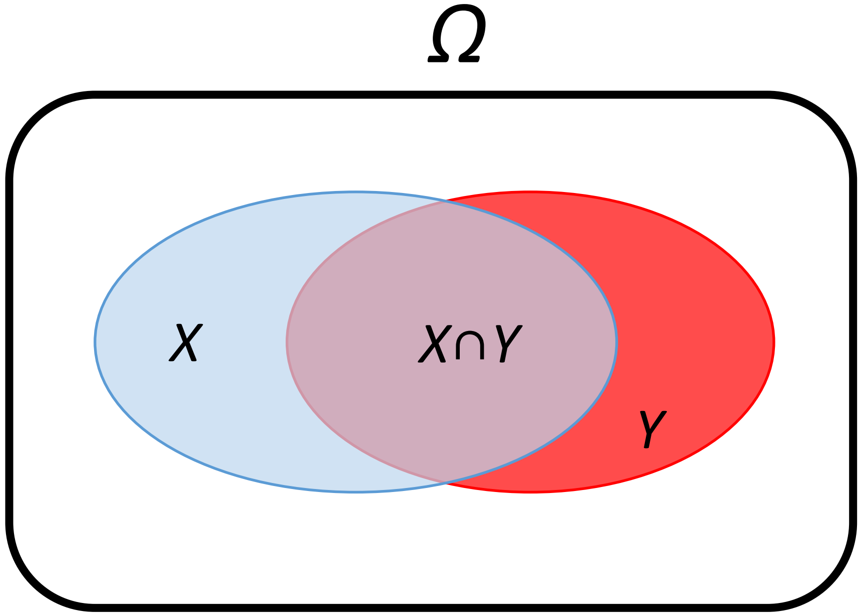
Example - solution

$$P\left(X < \frac{1}{2} \mid Y = y\right) = \int_0^{\frac{1}{2}} \frac{3x^2 + 3y^2 + 2xy}{3y^2 + y + 1} dx$$

$$= \frac{\frac{3}{2}y^2 + \frac{y}{4} + \frac{1}{8}}{3y^2 + y + 1}$$

As expected, depends on y

Bayes' Rule



Bayes' Rule relates the marginal and conditional probabilities of two events X and Y.

Theorem: Given events X and Y with $P(Y) > 0$:

$$P(X \mid Y) = \frac{P(Y \mid X)\, P(X)}{P(Y)}$$

Bayesian interpretation: X models an event that we cannot directly observe (i.e. whether a patient has a disease), while Y represents an observable event that provides information about X (i.e. whether a diagnostic test was positive).

Here:

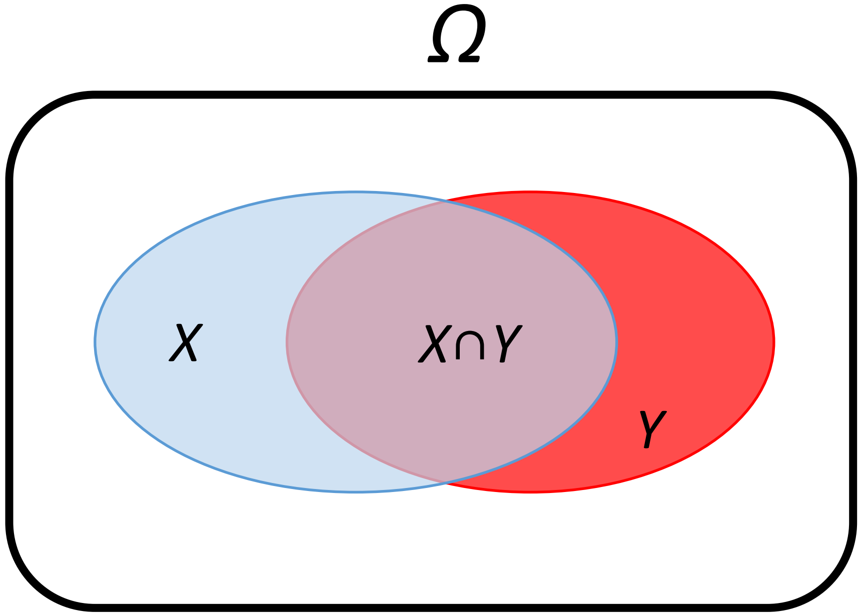
- $P(X)$ is called the *prior* probability of X. It represents our initial belief about proposition X being true.
- $P(Y \mid X)$ is called the *likelihood*. It models how the unobservable event X affects the observable event Y.
- $P(Y)$ is called the *evidence* – this is the *marginal* probability of observing Y (i.e. irrespective of X).
- $P(X \mid Y)$ is called the *posterior of X given Y*. It models our belief about X *after* incorporating information about Y.

Bayes' Rule

Bayes' Rule relates the marginal and conditional probabilities of two events X and Y.

Theorem: Given events X and Y with $P(Y) > 0$:

$$P(X \mid Y) = \frac{P(Y \mid X)\, P(X)}{P(Y)}$$



Main Idea:
When viewed in this way, Bayes' Rule provides a prescription for *updating our prior belief about X after observing data Y*.

Example application: Diagnostic testing

The probability that a person in a population has disease *D* is .1%. A test *T* for disease *D* has a false negative rate of 1% and a false positive rate of 5%. Given that a person receives a positive test result, what is the probability that they *actually* have the disease?

$$P(X \mid Y) = \frac{P(Y \mid X)\, P(X)}{P(Y)}$$

Solution: We want to calculate $P(D = 1 \mid T = 1)$

- **Prior:** $P(D = 1) = .001$
- **Likelihood:** $P(T = 0 \mid D = 1) = .01$, $P(T = 1 \mid D = 0) = .05$
- **Evidence calculation:** Using Law of Total Probability:

$$P(T = 1) = \sum_{k \in \{0,1\}} P(T = 1 \mid D = k) P(D = k)$$

$$= P(T = 1|D = 1)P(D = 1) + P(T = 1|D = 0)P(D = 0)$$

$$= (1 - .01)(.001) + (.05)(1 - .001) = .05094$$

- **Apply Bayes' Rule:**

$$P(D = 1 \mid T = 1) = \frac{P(T = 1 \mid D = 1)P(D = 1)}{P(T = 1)} = \frac{(1 - .01)(.001)}{.05094} = 0.01943$$

⇒ Even with a positive test result, the prob. a person actually has the disease is only ≈ 2%.

Bayes’ Rule in Robotics

A fundamental principle used in various applications

- **Localization:** The probability of being in a kitchen given a refrigerator is observed is higher that the probability of being in a bedroom given the observation of refrigerator.
- **Planning:** The probability of a robot’s next steps given the steps the robot has already executed.
- **Classification:** The probability of a new example is in a class given the training data set.
- **State estimation:** The probability of being in a particular state at current time step t given past actions and measurements.
- Many others…

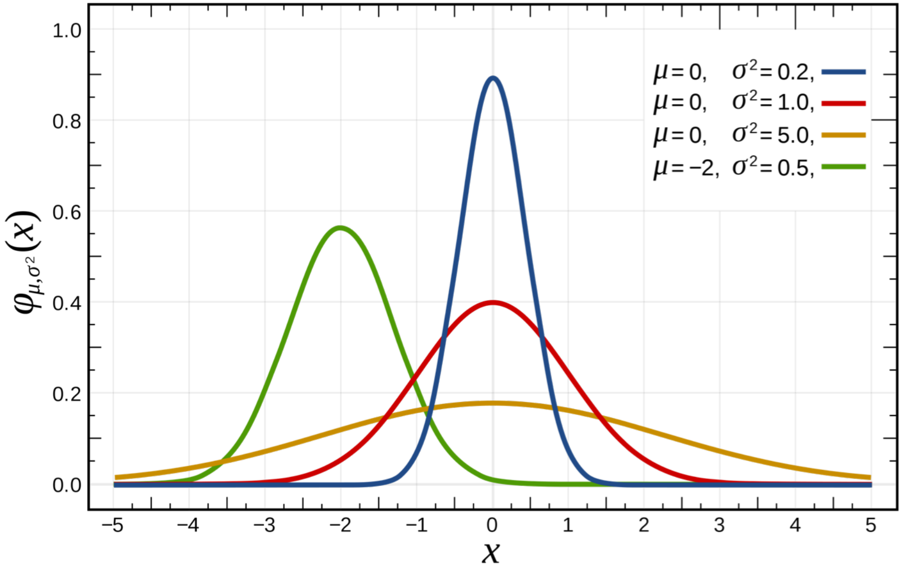
Normal distribution

A type of continuous probability distribution for a real-valued random variable

Given mean μ and standard deviation σ, its probability density function is

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The variance of the distribution is σ^2.



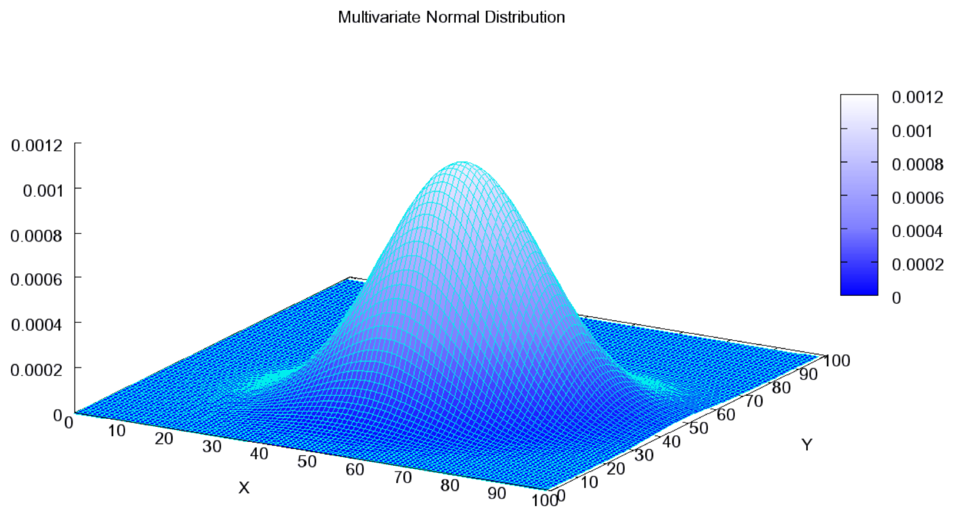
Multivariate Gaussian (Normal) distributions

Generalization of the one-dimensional normal distribution to higher dimensions

The *multivariate Gaussian distribution* with *mean μ* and *covariance Σ*, denoted $N(\mu, \Sigma)$, is the probability distribution on $\mathbb{R}^n$ determined by the density:

$$p(x \mid \mu, \Sigma) = (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

If k=1, it reduces to one-dimensional Normal dist.



Multivariate Gaussian (Normal) distributions

Key facts: Let $(X, Y) \sim N(\mu, \Sigma)$ where:

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}$$

Then:

- The marginal distribution for X is $N(\mu_X, \Sigma_{XX})$
- The conditional distribution for X given $Y = y$ is $N\left(\mu_{X|Y}, \Sigma_{X|Y}\right)$, where:

$$\mu_{X|Y} = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y), \qquad \Sigma_{X|Y} = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$$

- If $X \sim N(\mu, \Sigma)$ is a multivariate Gaussian and $Y = AX + b$ is its image under an *affine transformation*, then $Y \sim N\left(A\mu + b, A^{\mathrm{T}}\Sigma A\right)$

Covariance matrix:

$$\Sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$$

- Measures the direction of the relationship between two variables
- +: two variables tend to be high/low at the same time
- - : one is high, the other is low

Main takeaway: The family of multivariate Gaussians is *closed* under basic operations on prob. dist.
⇒ This is a super convenient model class!

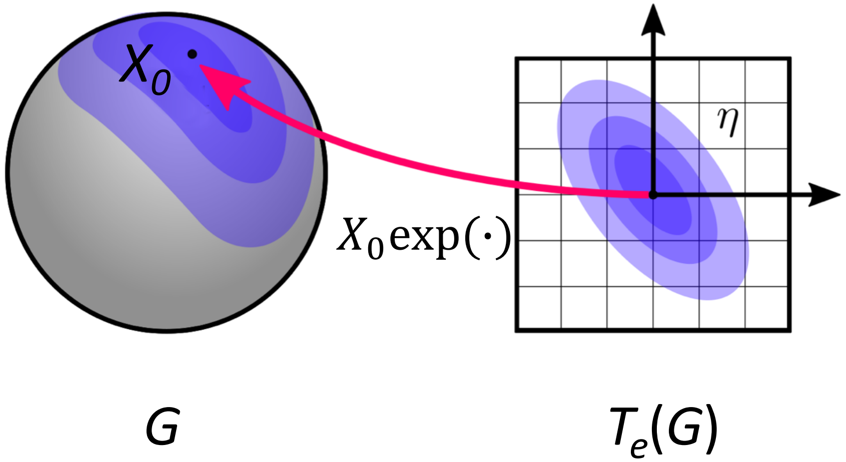
Modeling uncertainty on Lie groups

Let *G* be a Lie group of dimension *d*. Since $T_e(G) \cong \mathbb{R}^d$, and $\exp\colon T_e(G) \to G$, we can define a *G*-valued random variable *X* by:

$$\eta \sim N(0, \Sigma)$$
$$X = X_0 \exp(\eta)$$

where:

- The group element $X_0 \in G$ acts as a *location* parameter
- The covariance Σ is a *dispersion* parameter.



This gives a way of defining distributions on *arbitrary Lie groups*