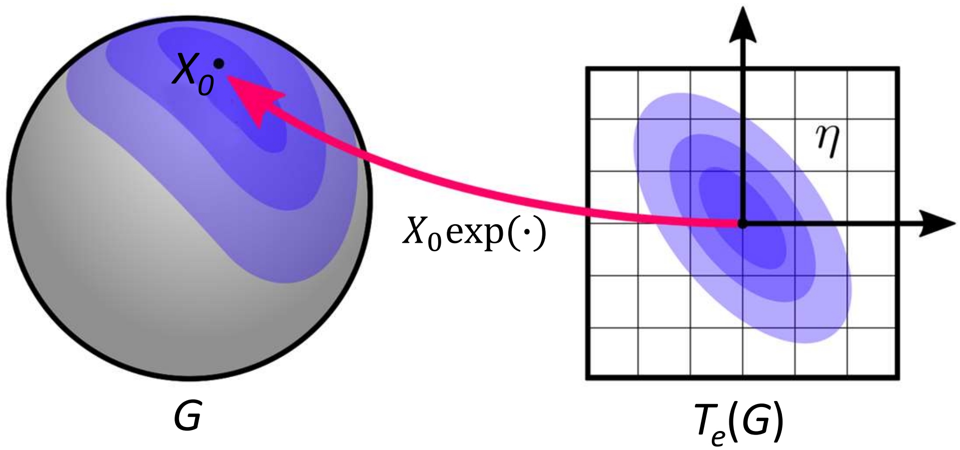
EECE 5550: Mobile Robotics



Lecture 4: Review of Probability

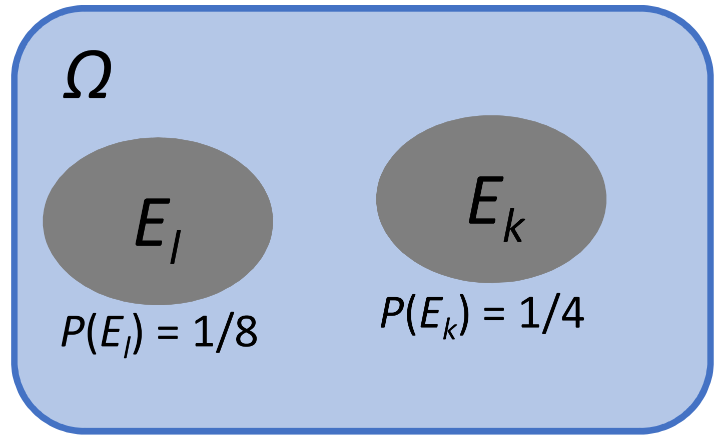
Plan of the day

- Review of probability theory (with a twist ☺)
- Uncertainty representation on Lie groups

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* $P\colon \Omega \rightarrow [0,1]$ that assigns a *probability* $P(E) \epsilon [0,1]$ to each event *E* in *F*.



$P(E_k \cup E_l) = 3/8$

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:

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

Intuitively:

- The sample space Ω is the set of *possible outcomes* of a random process
- An event E_k describes a *range* of possible outcomes
- The probability $P(E_k)$ describes the likelihood of a random outcome lying in the range E_k.

Why think about probability this way?

Reason #1: 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*

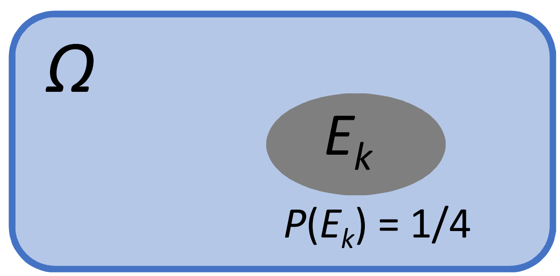
Main point: *Both* of these approaches are just ways of *measuring the "size" of a set*

Continuous Case

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

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

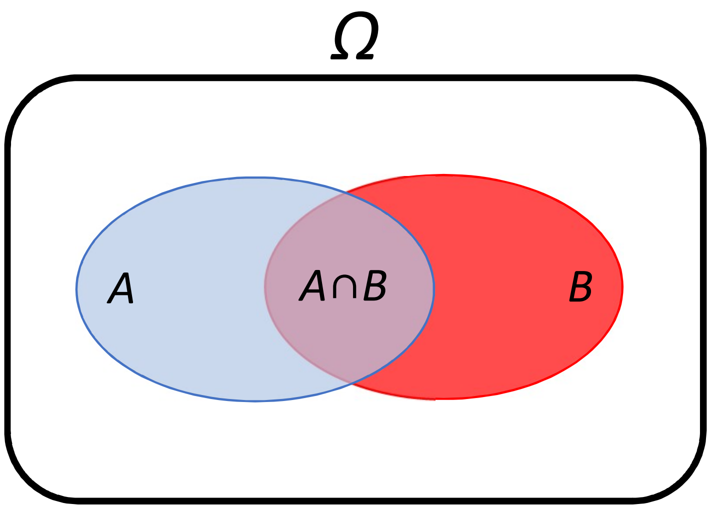
- Defining object: *probability density function*



Why think about probability this way?

Reason #2: It implies that statements about "probability" are really just statements about *sizes of sets*

⇒ "Proof by Venn diagram"



Examples / useful identities:

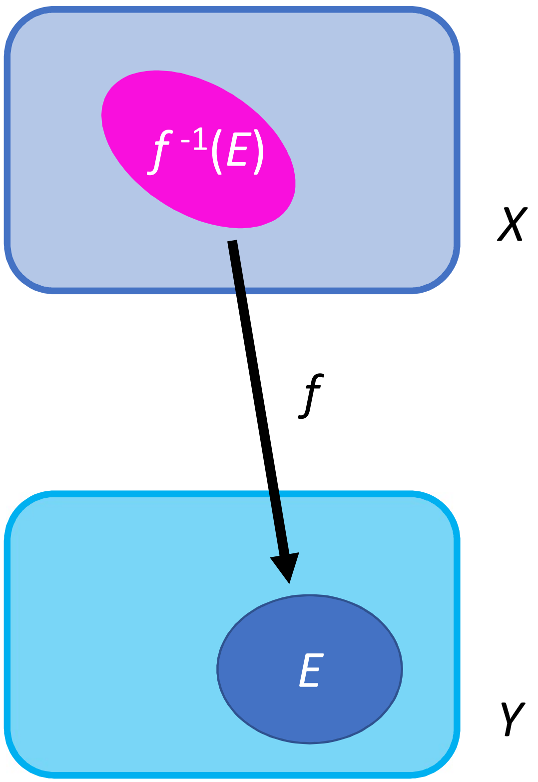
- **Complement rule:** $P(A) + P(A^c) = 1$
- **Sum rule:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Union bound:** $P(A \cup B) \leq P(A) + P(B)$
- **Law of total probability:** If $\{B_K\}$ is a *countable partition* of Ω, then:

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

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*.

Examples:

- **Random variables:** Formally, a *random variable* *V* is a *function* $V : X \to Y$ from a probability space *X* to some sample space *Y*. For $E \subseteq Y$, we have $P(V \in E) = f_* P(E)$.

  NB: This is very general! *Y* can be a set of *objects*, e.g. graphs, etc.

- **Special case: Change of variables:** If *V* is a random variable on $\Omega \subseteq \mathbb{R}^n$ with density p_V and $f : \Omega \to Y \subseteq \mathbb{R}^m$, then $W \triangleq f(V)$ is a random variable on *Y* with probability density:

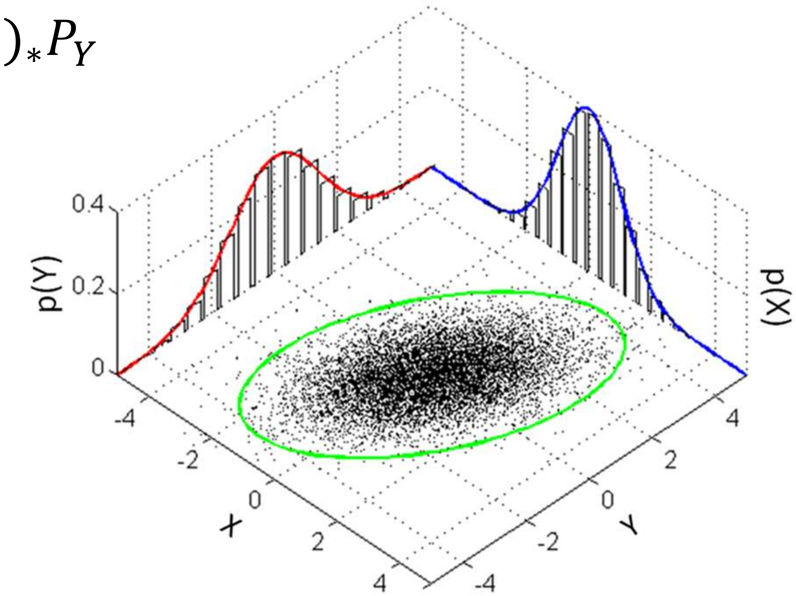
$$p_W(y) = \sum_{x \in f^{-1}(y)} p_V(x) \left| \frac{df}{dx}(x) \right|^{-1}$$

Joint and marginal distributions

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

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

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



Intuitively:

- P_{XY} is a (*joint*) distribution over *pairs* $(x, y) \in X \times Y$
- The marginals P_X and P_Y represent the induced distributions over *x* and *y* *alone* (i.e., ignoring the elements *y* and *x* in the pair), respectively.

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:

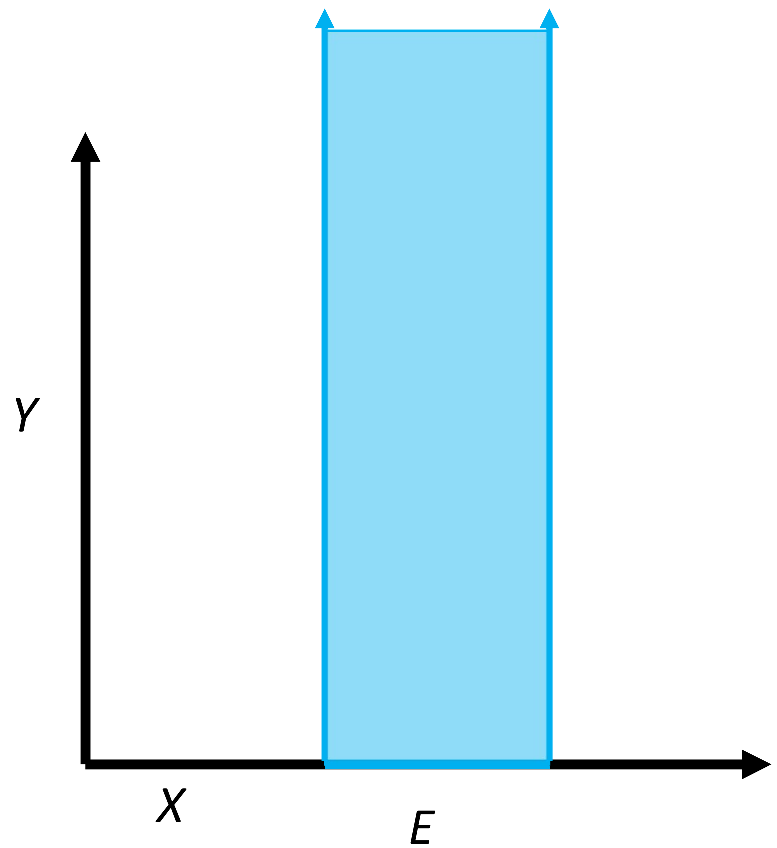
$$p_X(x) = \int_{\mathbb{R}^n} p_{XY}(x, y) dy$$

Joint and marginal distributions

Derivation of marginal density:

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

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

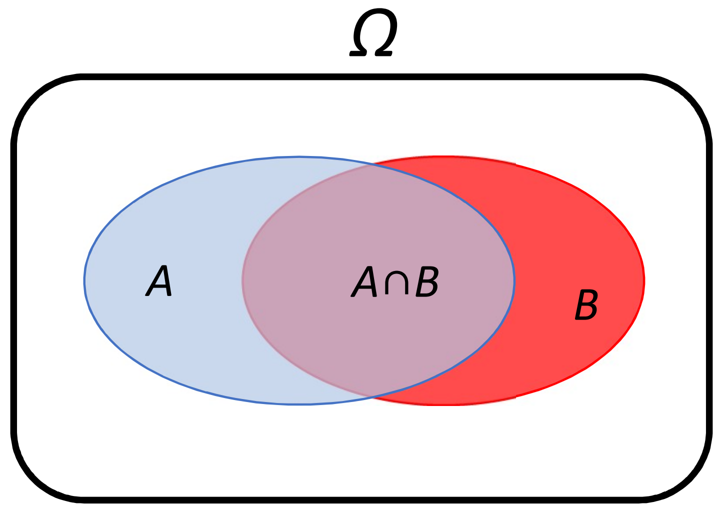


Conditional probability and independence

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

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

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



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, $P(A \cap B) = P(A)P(B)$.

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

Derivation of conditional density:

$$P(X \in A \,|Y \in B) = \frac{P((X \in A) \cap (Y \in B))}{P(Y \in B)}$$

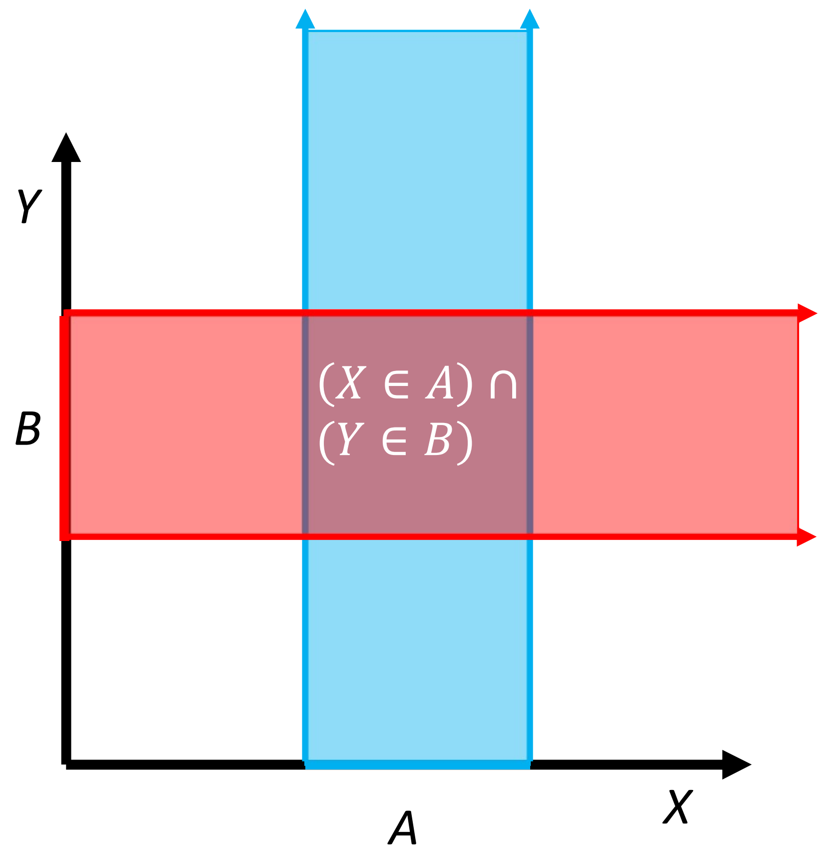
$$= \frac{1}{P(Y \in B)} \int_{A \times B} p_{XY}(x, y)\, dx\, dy$$

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

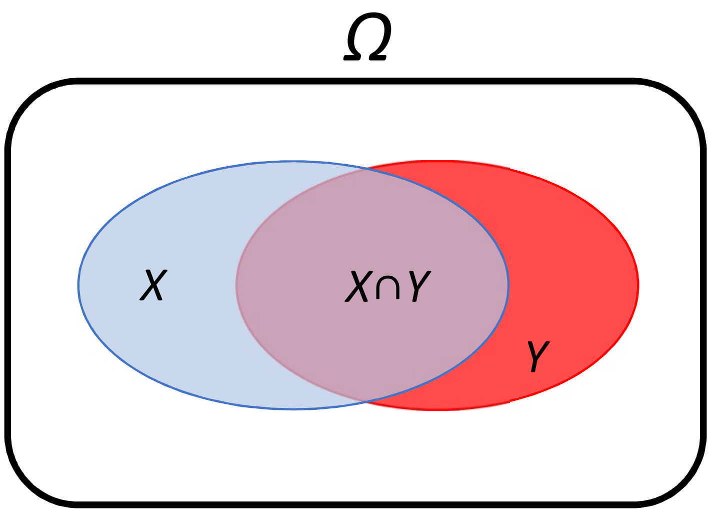
$$\Rightarrow \; p_{X|Y}(x \mid Y \in B) = \frac{1}{P(Y \in B)} \int_B p_{XY}(x, y)\, dy$$

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

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



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* | *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* | *Y*) is called the *posterior* of *X* given *Y*. It models our belief about *X* *after* incorporating information about *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:

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

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

$$= (1 - .01)(.001) + (.05)(1\text{-}.001) = .05094$$

- **Apply Bayes' Rule:**

$$P(D = 1 \mid T = 1) = \frac{P(T = 1 \mid D = 1) P(D = 1)}{P(T = 1)} = \frac{(1 - .01)(.001)}{.05094} = 0.01093$$

$\Rightarrow$ Even with a positive test result, the probability a person actually has the disease is only $\approx 1\%$.

Multivariate Gaussian distributions

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:

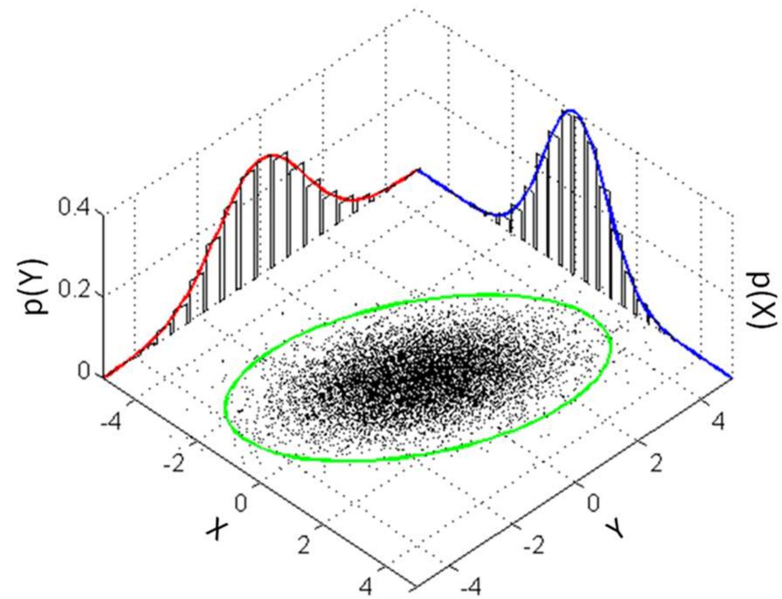
$$p(x \mid \mu, \Sigma) = (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma (x-\mu)}$$

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 $N(\mu_X, \Sigma_{\mathrm{XX}})$
- 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 $Y \sim N\left(A\mu + b, A^{\mathrm{T}}\Sigma A\right)$



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

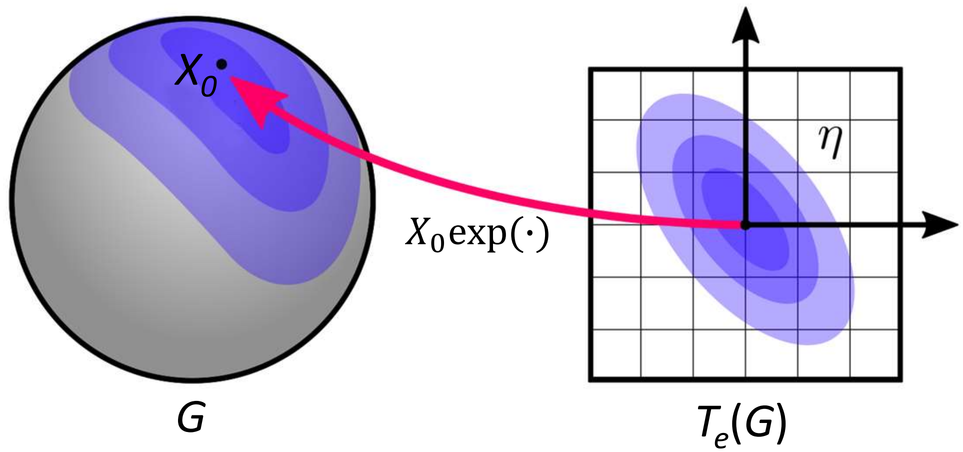
Distributions on Lie groups

Let G be a Lie group of dimension d. Since $T_e(G) \cong \mathbb{R}^d$, and $\exp: T_e(G) \to G$, 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.



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