

# Probability: A Second Dose in ROB 501

## 1 Probability Spaces

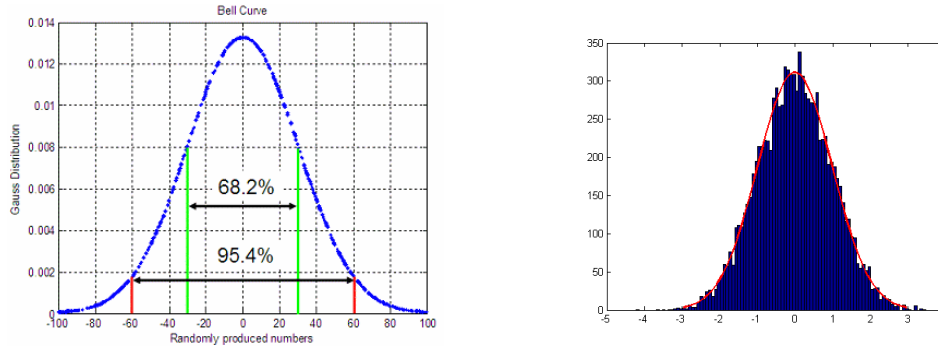


Figure 1: (Left) Normal distribution  $N(\mu, \sigma)$  with  $\mu = 0$  and  $\sigma = 30$ . (Right) How do you determine the density? You have to collect data! The figure shows a “fit” of a normal distribution to data.

**Def.**  $(\Omega, \mathcal{F}, P)$  is called a probability space.

- $\Omega$  is the sample space. Think of it as the domain of a random variable  $X : \Omega \rightarrow \mathbb{R}$  or random vector  $X : \Omega \rightarrow \mathbb{R}^m$ .
- $A \subset \Omega$  is an event.
- $\mathcal{F}$  is the collection of allowed events<sup>1</sup>. It must at least contain  $\emptyset$  and  $\Omega$ . It is closed w.r.t. countable unions and intersections, and set complement.
- $P : \mathcal{F} \rightarrow [0, 1]$  is a probability measure. It has to satisfy a few basic operations
  - $P(\emptyset) = 0$  and  $P(\Omega) = 1$ .
  - For each  $A \in \mathcal{F}$ ,  $0 \leq P(A) \leq 1$
  - If the sets  $A_1, A_2, \dots$  are disjoint (i.e.,  $A_i \cap A_j = \emptyset$  for  $i \neq j$ ), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- Example event  $A := \{\omega \in \Omega \mid |\omega - \mu| \leq \sigma\} \Rightarrow P(A) = 0.682$

<sup>1</sup>Though it is too deep for ROB 501, there are subsets of the reals, for example, that are so complicated one cannot define a reasonable notion of probability that agrees with how we would want to define the probability of an interval, such as  $[a, b]$ .

## 2 Random Variables

**Def.**  $X : \Omega \rightarrow \mathbb{R}$  is a random variable if  $\forall x \in \mathbb{R}$ , the set  $\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$ . This just means that such sets can be assigned probabilities.

**Remarks:**

- Shorthand notation  $\{X \leq x\} := \{\omega \in \Omega \mid X(\omega) \leq x\}$
- Because  $\mathcal{F}$  is closed under set complements, (countable) unions, and (countable) intersections, we can also assign probabilities to
  - a)  $\{X > x\} = \sim \{X \leq x\} = \{X \leq x\}^C$
  - b)  $\{x < X \leq y\} = \{X \leq y\} \cap \{X > x\}$

**Def.**  $X : \Omega \rightarrow \mathbb{R}$  is a continuous random variable if there exists a density  $f : \mathbb{R}^p \rightarrow [0, \infty)$  such that,

$$\forall x \in \mathbb{R}, \quad P(\{X \leq x\}) = \int_{-\infty}^x f(\bar{x}) d\bar{x} \quad (\bar{x} \text{ dummy variable in the integral})$$

**Remarks:**

- $\int_a^b f(x) dx = P(a < X \leq b) = P(a \leq X \leq b)$   
 $= P(\{\omega \in \Omega \mid X(\omega) \in [a, b]\})$
- mean:  $\mu := \mathcal{E}\{X\} := \int_{-\infty}^{\infty} x f(x) dx$
- Variance:  $\sigma^2 := \mathcal{E}\{(X - \mu)^2\} := \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$
- Standard Deviation  $\sigma := \sqrt{\sigma^2}$  (Std. Dev.)

### 3 Random Vectors

**Def.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space. A function  $X : \Omega \rightarrow \mathbb{R}^p$  is called a random vector if each component of  $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$  is a random variable, that is,  $\forall 1 \leq i \leq p$ ,  $X_i : \Omega \rightarrow \mathbb{R}$  is a random variable.

**Consequence:**  $\forall x \in \mathbb{R}^p$ , the set  $\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$  (i.e., it is an allowed event), where the inequality is understood pointwise, that is,

$$\{\omega \in \Omega \mid X(\omega) \leq x\} := \{\omega \in \Omega \mid \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_p(\omega) \end{bmatrix} \leq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}\} = \bigcap_{i=1}^p \{\omega \in \Omega \mid X_i(\omega) \leq x_i\}$$

**Def.**  $X : \Omega \rightarrow \mathbb{R}^p$  is a continuous random vector if there exists a density  $f : \mathbb{R}^p \rightarrow [0, \infty)$  such that,

$$\forall x \in \mathbb{R}^p, \quad P(\{X \leq x\}) = \int_{-\infty}^{x_p} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p) d\bar{x}_1 d\bar{x}_2 \dots d\bar{x}_p$$

## 4 Moments

**Def.** Suppose  $g : \mathbb{R}^p \rightarrow R$

$$E\{g(X)\} = \int_{\mathbb{R}^p} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_p) f_X(x_1, \dots, x_p) dx_1 \dots dx_p$$

### Mean or Expected Value

$$\mu = E\{X\} = E\left\{\begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}\right\} = \begin{bmatrix} \mathcal{E}\{X_1\} \\ \vdots \\ \mathcal{E}\{X_p\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}$$

### Covariance Matrices

$$\Sigma := \text{cov}(X) = \text{cov}(X, X) = E\{(X - \mu)(X - \mu)^T\}$$

where

$$(X - \mu) \text{ is } p \times 1, (X - \mu)^T \text{ is } 1 \times p, (X - \mu)(X - \mu)^T \text{ is } p \times p$$

**Exercise**  $\text{cov}(X)$  is positive semidefinite

## 5 Marginal Densities, Independence, and Correlation

Suppose the random vector  $X : \Omega \rightarrow \mathbb{R}^p$  is partitioned into two components  $X_1 : \Omega \rightarrow \mathbb{R}^n$  and  $X_2 : \Omega \rightarrow \mathbb{R}^m$ , with  $p = n + m$ , so that,

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

**Notation:** We denote the density of  $X$  by

$$f_X(x) = f_{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}}(x_1, x_2) = f_{X_1 X_2}(x_1, x_2)$$

and it is called the joint density of  $X_1$  and  $X_2$ . As before, we can define the mean and covariance.

- Mean is  $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mathcal{E}\{X\} = E\left\{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}\right\} = \begin{bmatrix} \mathcal{E}\{X_1\} \\ \mathcal{E}\{X_2\} \end{bmatrix}$

- Covariance is

$$\begin{aligned} \Sigma &= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \mathcal{E}\left\{\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}^\top\right\} \\ &= \mathcal{E}\left\{\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} [(X_1 - \mu_1)^\top \quad (X_2 - \mu_2)^\top]\right\} \\ &= \mathcal{E}\left\{\begin{bmatrix} (X_1 - \mu_1)(X_1 - \mu_1)^\top & (X_1 - \mu_1)(X_2 - \mu_2)^\top \\ (X_1 - \mu_1)(X_2 - \mu_2)^\top & (X_2 - \mu_1)(X_2 - \mu_2)^\top \end{bmatrix}\right\} \end{aligned}$$

where  $\Sigma_{12} = \Sigma_{21}^\top = \text{cov}(X_1, X_2) = \mathcal{E}\{(X_1 - \mu_1)(X_2 - \mu_2)^\top\}$  is also called the correlation of  $X_1$  and  $X_2$ .

If  $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} : \Omega \rightarrow \mathbb{R}^{n+p}$  is a continuous random vector, then its components

$$X_1 : \Omega \rightarrow \mathbb{R}^n \quad \text{and} \quad X_2 : \Omega \rightarrow \mathbb{R}^m$$

are also continuous random vectors and have densities,  $f_{X_1}(x_1)$  and  $f_{X_2}(x_2)$ . These densities are given a special name.

**Def.**  $f_{X_1}(x_1)$  and  $f_{X_2}(x_2)$  are called the marginal densities of  $X_1$  and  $X_2$ .

**Fact:** In general the marginal densities are a nightmare to compute.

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2) dx_2 \\ &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2}(\underbrace{\bar{x}_1, \dots, \bar{x}_n}_{x_1}, \underbrace{\bar{x}_{n+1}, \dots, \bar{x}_{n+m}}_{x_2}) \underbrace{d\bar{x}_{n+1} \cdots d\bar{x}_{n+m}}_{dx_2} \end{aligned}$$

$$\begin{aligned} f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2) dx_1 \\ &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 X_2}(\underbrace{\bar{x}_1, \dots, \bar{x}_n}_{x_1}, \underbrace{\bar{x}_{n+1}, \dots, \bar{x}_{n+m}}_{x_2}) \underbrace{d\bar{x}_1 \cdots d\bar{x}_n}_{dx_1} \end{aligned}$$

For Normal Random Vectors, however, we can read them directly from the joint density! We will not be doing any iterated integrals.

**Def.**  $X_1$  and  $X_2$  are independent random vectors if their joint density factors

$$f_X(x) = f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

**Def.**  $X_1$  and  $X_2$  are uncorrelated if their “cross covariance” or “correlation ” is zero

$$\text{cov}(X_1, X_2) := \mathcal{E}\{(X_1 - \mu_1)(X_2 - \mu_2)^\top\} = 0_{n \times m}$$

**Fact:** If  $X_1$  and  $X_2$  are independent, then they are also uncorrelated. **The converse is in general false.**

## 6 Conditioning

**Def.** For two events  $A, B \in \mathcal{F}, P(B) > 0$

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

is the conditional probability of  $A$  given  $B$ .

**Remarks:**

- Suppose  $P(A)$  is our current estimate of the probability that our robot is near a certain location and  $B$  is a measurement of the robot's location, with confidence in the measurement being  $P(B)$ . The conditional probability of  $A$  given  $B$  occurred is how we “fuse” the two pieces of information

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

- 

$$B \subset A, P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

- 

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

Consider again our partitioned random vector  $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

**Def.** The conditional density of  $X_1$  given  $X_2 = x_2$  is

$$f_{X_1|X_2}(x_1 | x_2) = \frac{f_{X_1X_2}(x_1, x_2)}{f_{X_2}(x_2)}.$$

Sometimes we simply write  $f(x_1 | x_2)$

### Remarks on Conditional Random Vectors:

- **Very important:**  $X_1$  given  $X_2 = x_2$  is (still) a random vector. It's density is  $f_{X_1|X_2}(x_1 | x_2)$
- **Conditional Mean:**

$$\mu_{X_1|X_2=x_2} := \mathcal{E}\{X_1 | X_2 = x_2\}$$

$$:= \int_{-\infty}^{\infty} x_1 f_{X_1|X_2}(x_1 | x_2) dx_1$$

$\mu_{X_1|X_2=x_2}$  is a function of  $x_2$ . Think of it as a function of the value read by your sensor!

- **Conditional Covariance:**

$$\Sigma_{X_1|X_2=x_2} := \mathcal{E}\{(X_1 - \mu_{X_1|X_2=x_2})(X_1 - \mu_{X_1|X_2=x_2})^\top | X_2 = x_2\}$$

$$:= \int_{-\infty}^{\infty} (X_1 - \mu_{X_1|X_2=x_2})(X_1 - \mu_{X_1|X_2=x_2})^\top f_{X_1|X_2}(x_1 | x_2) dx_1$$

$\Sigma_{X_1|X_2=x_2}$  is a function of  $x_2$ . Think of it as a function of the value read by your sensor!