Statistics for Chemical Engineers:
From Data to Models to Decisions

Victor M. Zavala

Department of Chemical and Biological Engineering
University of Wisconsin-Madison

victor.zavala@wisc.edu

Chapter 3: Multivariate Random Variables

Multivariate Random Variables

- We have assumed X is univ and thus realization $\omega \in \Omega$ generates a scalar observation $x_\omega \in \mathbb{R}$. This implies $\mathcal{D}$ is unidim.
- Consider multiv $X = (X_1, X_2, ..., X_n)$, in which realization $\omega \in \Omega$ generates an observation vector $x_\omega = (x_{\omega,1}, x_{\omega,2}, ..., x_{\omega,n}) \in \mathbb{R}^n$. This implies $\mathcal{D}$ is multidim.

Questions that we are interested in answering are:

- Are there any connections between RVs? Is there a clear trend that suggests they vary together? Do they vary independently of one another?
- How strong are connections?
- How does knowledge of one resolves uncertainty of the other?
- How to analyze connections between many RVs? (e.g., n is in the hundreds)?

Multivariate Random Variables

These questions are relevant from a *decision-making* standpoint:

- If data suggests that RVs are related, is there a hidden *mechanistic* relationship?
- Can we exploit such relationships to develop *predictive* models and make decisions?

Elements of multivariate RV are similar to those of univariate RV but dependence and independence introduce a number of new concepts.

As in univariate case, our goal will also be to understand if our data seems to follow the behavior of a multivariate RV model.

Unfortunately, there are extremely few multivariate RV models that are of practical use. So most multivariate analysis is data-driven, but this provides key insights.

Domains and Subdomains

- Multivariate RVs give rise to data and functions that live in multiple dimensions.
- Domains of multiv RVs can take complex shapes (e.g., cubes, ellipses, polyhedra).
- Understanding domains helps visualize and analyze data over high dimensions.
- Understanding domains also helps understand events and decision-making logic.
- For simplicity, we focus on $X = (X_1, X_2)$.
- Here, $\omega \in \Omega$ of $X = (X_1, X_2)$ generates pair $x_\omega = (x_{\omega,1}, x_{\omega,2}) \in \mathbb{R}^2$.
- For example, a reactor experiment generates data on pressure and conversion.

Domains and Subdomains

- Domain of $X = (X_1, X_2)$ is $\mathcal{D} \subseteq \mathbb{R}^2$ and results from:

$$\mathcal{D} = \mathcal{D}_{X_1} \times \mathcal{D}_{X_2}$$

  where $\mathcal{D}_{X_1} \subseteq \mathbb{R}$ is domain of the X_1 and $\mathcal{D}_{X_2} \subseteq \mathbb{R}$ is domain of X_2.

- For *discrete* X, $\mathcal{D}_1$ and $\mathcal{D}_2$ are discrete. If $\mathcal{D}_1 = \{l_1, u_1\}$ and $\mathcal{D}_2 = \{l_2, u_2\}$, then:

$$\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 = \{(l_1, l_2), (l_1, u_2), (u_1, l_2), (u_2, l_2)\}.$$

- For *continuous* X, $\mathcal{D}_1$ and $\mathcal{D}_2$ are continuous. If $\mathcal{D}_1 = [l_1, u_1]$ and $\mathcal{D}_2 = [l_2, u_2]$:

$$\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 = [l_1, u_1] \times [l_2, u_2].$$

Domains and Subdomains

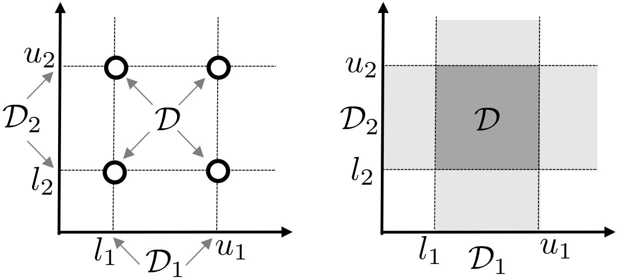


Figure: Domains for 2D discrete (left) and continuous (right) RV.

Domains and Subdomains

- *Empirical approx* of $\mathcal{D}$ is $\hat{\mathcal{D}}$, this is region occupied by available data $x_\omega,\ \omega \in \mathcal{S}$.
- Many observations might be needed to cover multidimensional domain.
- For example, in 3D, domain of $X = (X_1, X_2, X_3)$ can be a cube.
- Large portions of actual domain might remain unobserved.
- Important to have model to know what is possible beyond available data.

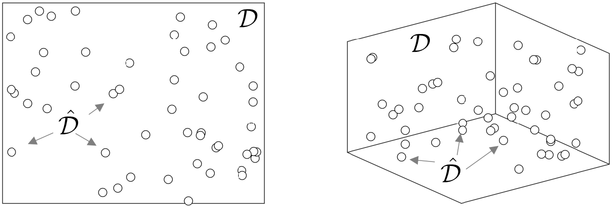


Figure: Empirical (observed) and theoretical 2D and 3D domains.

Domains and Subdomains

- Subdomains $\mathcal{A} \subseteq \mathcal{D}$ define *events* associated with X.
- For example, $\mathbb{P}(X \in \mathcal{A})$ is probability of event $X \in \mathcal{A}$.
- Imagine we want probability of event $X \in \mathcal{R} = \{X_1 \geq 0 \,\&\, X_2 \geq 0\}$ (e.g., prob that X takes non-negative values).
- To construct subdomain $\mathcal{A} \subseteq \mathcal{D}$ that captures event defined by $\mathcal{R}$, we note that any observation $x_\omega \in \mathcal{A}$ must satisfy $x_\omega \in \mathcal{R}$ *and* $x_\omega \in \mathcal{D}$. Consequently:

$$
\begin{aligned}
\mathcal{A} &= \mathcal{R} \cap \mathcal{D} \\
&= \{\{X_1 \geq 0 \,\&\, x_2 \geq 0\} \;\&\; \{l_1 \leq X_1 \leq u_1 \;\&\; l_2 \leq X_2 \leq u_2\}\} \\
&= \{X_1 \geq 0 \;\&\; X_2 \geq 0 \;\&\; l_1 \leq X_1 \leq u_1 \;\&\; l_2 \leq X_2 \leq u_2\}.
\end{aligned}
$$

  Symbol "$\cap$" denotes *set intersection* and captures "$\&$" logic

- In other words, event $\mathcal{A}$ must satisfy all the restrictions imposed by $\mathcal{R}$ and $\mathcal{D}$.
- This means domain $\mathcal{A}$ is more restrictive than domain $\mathcal{D}$ and thus $\mathcal{A} \subseteq \mathcal{D}$.

Domains and Subdomains

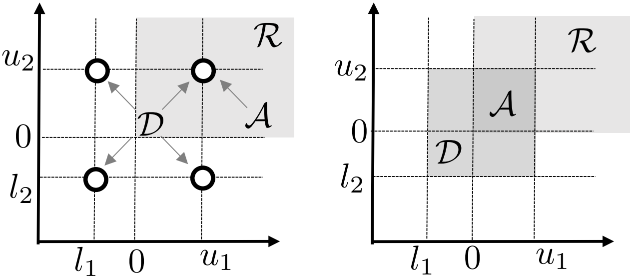


Figure: Subdomain $\mathcal{A}$ of $\mathcal{D}$ for discrete (left) and continuous (right) RV.

Domains and Subdomains

- Different types of events $\mathcal{A}$ capture different *logic*.
- For instance, we can build subdomain $\mathcal{A}$ that enforces constraint that $\{X_1 = X_2\}$ (to compute probability that $X_1 = X_2$).
- Can also build a subdomain $\mathcal{A}$ that enforces constraint that $\{X_1^2 + X_2^2 \leq 1\}$ (to compute probability that X is in a circular region).

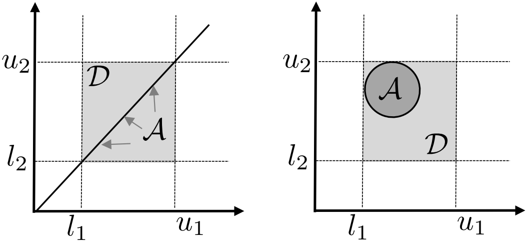


Figure: Examples of subdomains $\mathcal{A} \subseteq \mathcal{D}$ capturing different constraints.

Domains and Subdomains

- When making decisions, we are interested in events that capture "*or*" logic.
- Want prob that X is in $\mathcal{R}_1 = \{X_1 \geq 0 \,\&\, X_2 \geq 0\}$ *or* that it is in $\mathcal{R}_2 = \{X_1 \leq -1 \,\&\, X_2 \leq -1\}$.
- Build subdomain $\mathcal{A}$ by defining $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$, where $\cup$ is *set union*.
- Here, $\mathbb{P}(X \in \mathcal{R})$ indicates probability that $X \in \mathcal{R}_1$ or that $X \in \mathcal{R}_2$.
- $\mathcal{A}$ must be a subdomain of $\mathcal{D}$, we achieve this by enforcing $\mathcal{A} = \mathcal{R} \cap \mathcal{D}$ (X is in $\mathcal{R}$ and $\mathcal{D}$).

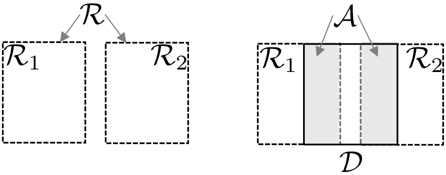


Figure: Union of $\mathcal{R}_1$ and $\mathcal{R}_2$ to obtain $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2$ (left). Intersection of $\mathcal{R}$ with $\mathcal{D}$ to obtain event $\mathcal{A} = \mathcal{R} \cap \mathcal{D}$ (right).

Domains and Subdomains

- In decision-making, we also often want probability that $\mathcal{A}_1$ occurs given knowledge that $\mathcal{A}_2$ has occurred.
- This is captured by *conditional probability* $\mathbb{P}(\mathcal{A}_1|\mathcal{A}_2)$:

$$\mathbb{P}(\mathcal{A}_1|\mathcal{A}_2) = \frac{\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2)}{\mathbb{P}(\mathcal{A}_2)}.$$

- $\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2)$ is *joint* prob and $\mathbb{P}(\mathcal{A}_2)$ is marginal prob of $\mathcal{A}_2$ (similar for $\mathcal{A}_1$).
- Expression also written as $\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) = \mathbb{P}(\mathcal{A}_1|\mathcal{A}_2) \cdot \mathbb{P}(\mathcal{A}_2)$.
- Intuitively, if $\mathcal{A}_2$ has occurred and overlap $\mathcal{A}_1 \cap \mathcal{A}_2$ is large, then there is a high probability that $\mathcal{A}_1$ also occurs.
- If event $\mathcal{A}_2$ does not affect occurrence of $\mathcal{A}_1$ (events are *independent*) then

$$\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) = \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2)$$

- Disjoint events have an empty overlap $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$.

Domains and Subdomains

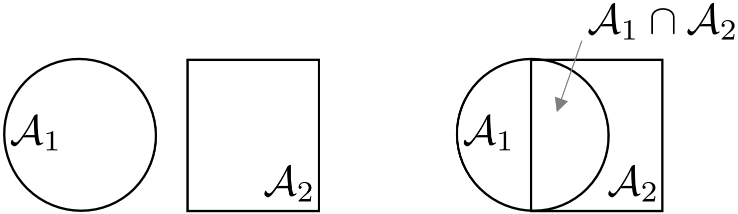


Figure: Overlapping (joint) events (right) and non-overlapping (disjoint) events (left).

Domains and Subdomains

- One can use the facts that:

$$\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) = \mathbb{P}(\mathcal{A}_1|\mathcal{A}_2) \cdot \mathbb{P}(\mathcal{A}_2)$$
$$\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) = \mathbb{P}(\mathcal{A}_2|\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_1)$$

  to establish that:

$$\mathbb{P}(\mathcal{A}_1|\mathcal{A}_2) = \frac{\mathbb{P}(\mathcal{A}_2|\mathcal{A}_1)\mathbb{P}(\mathcal{A}_1)}{\mathbb{P}(\mathcal{A}_2)}.$$

- This is an important result in statistics known as *Bayes' Rule*.

Joint PDFs and CDFs

- *Joint pdf* is used to quantify prob of events associated with $X = (X_1, X_2)$:

$$f(x_1, x_2) = \mathbb{P}(X_1 = x_1 \,\&\, X_2 = x_2), \;\; (x_1, x_2) \in \mathcal{D}.$$

- For discrete X, this is interpreted as probability of joint event in which X_1 takes value x_1 *and* X_2 takes value x_2.

- For continuous X, this is interpreted as probability that (X_1, X_2) is in an infinitesimal neighborhood around (x_1, x_2).

- Joint pdf for continuous X must satisfy:

$$f(x_1, x_2) \geq 0, \; (x_1, x_2) \in \mathcal{D}$$
$$\int_{(x_1, x_2) \in \mathcal{D}} f(x_1, x_2) dx_1 dx_2 = 1.$$

For discrete RV we replace integral with double summation.

Joint PDFs and CDFs

- Joint pdf used to compute probability of events $\mathcal{A} \subseteq \mathcal{D}$. For a continuous RV:

$$\mathbb{P}(X \in \mathcal{A}) = \int_{(x_1,x_2)\in\mathcal{A}} f(x_1,x_2)dx_1dx_2.$$

- Joint pdf is *density* of points at a point in subdomain $\mathcal{A}$.
- Probability $\mathbb{P}(X \in \mathcal{A})$ is *total density* of points in $\mathcal{A}$.
- This $\mathbb{P}(X \in \mathcal{A})$ measures size (in terms of density) of subdomain $\mathcal{A}$; i.e., the larger the subdomain, the higher the probability of being in it.

Joint PDFs and CDFs

- We are often interested in events associated with *box* subdomains $\mathcal{A} \subseteq \mathcal{D}$:

$$\begin{aligned}\mathcal{A} &= \{a_1 \leq x_1 \leq b_1 \ \&\ a_2 \leq x_2 \leq b_2\} \\ &= \{x_1 \in [a_1, b_1] \ \&\ x_2 \in [a_2, b_2]\}.\end{aligned}$$

- For continuous RVs, probability of being in such a box is:

$$\begin{aligned}\mathbb{P}(X \in \mathcal{A}) &= \int_{(x_1, x_2) \in \mathcal{A}} f(x_1, x_2) dx_1 dx_2 \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_1 dx_2.\end{aligned}$$

Empirical Event Probabilities

- The probability of an event is computed by using the joint pdf; for a continuous RV:

$$\mathbb{P}(X \in \mathcal{A}) = \int_{x \in \mathcal{A}} f(x)dx.$$

- In many practical situations, we do not have the joint pdf, but all we have available is data x_ω, $\omega \in \mathcal{S}$.

- It is rather straightforward to compute empirical estimates for all pdfs, cdfs, and probabilities discussed by using data. For instance:

$$\hat{\mathbb{P}}(X \in \mathcal{A}) = \frac{1}{S} \sum_{\omega \in \mathcal{S}} \mathbf{1}[x_\omega \in \mathcal{A}]$$

$$\hat{\mathbb{P}}(X \in \mathcal{A}_1 \cap \mathcal{A}_2) = \frac{1}{S} \sum_{\omega \in \mathcal{S}} \mathbf{1}[x_\omega \in \mathcal{A}_1 \ \& \ x_\omega \in \mathcal{A}_2]$$

$$\hat{\mathbb{P}}(X \in \mathcal{A}_1 \cup \mathcal{A}_2) = \frac{1}{S} \sum_{\omega \in \mathcal{S}} \mathbf{1}[x_\omega \in \mathcal{A}_1 \text{ or } x_\omega \in \mathcal{A}_2]$$

$$\hat{\mathbb{P}}(X \in \mathcal{A}_1 | X \in \mathcal{A}_2) = \frac{\hat{\mathbb{P}}(X \in \mathcal{A}_1 \ \& \ X \in \mathcal{A}_2)}{\hat{\mathbb{P}}(X \in \mathcal{A}_2)}.$$

Example: Lifetime of Thermostat

- Thermostat relies on sensor (thermocouple) and computing processor that processes sensor data and sends feedback signals to air conditioning unit.

- Thermostat fails to operate if any of these components fail.

- Time to failure of components is (X_1, X_2) has joint pdf:

$$f(x_1, x_2) = \frac{1}{\mu_1 \cdot \mu_2} e^{-(x_1/\mu_1 + x_2/\mu_2)}$$

with $\mathcal{D} = \{0 \leq x_1 \leq \infty \;\&\; 0 \leq x_2 \leq \infty\}$ and $\mu_1 = 1$ and $\mu_2 = 10$ years.

- Compute probability that both components last less than 1 year.
- Compute probability that of event that $X_1 \geq 1$ and $X_2 \geq 2$.

Example: Lifetime of Thermostat

- Probability that both components last less than 1 year corresponds to $\mathcal{A} = \{0 \leq x_1 \leq 1 \ \& \ 0 \leq x_2 \leq 1\}$.

- Clearly, $\mathcal{A} \subseteq \mathcal{D}$ and probability is:

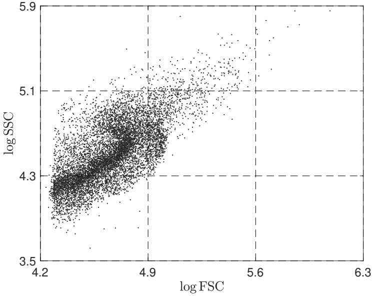
$$\begin{aligned}\mathbb{P}(X \in \mathcal{A}) &= \int_0^1 \int_0^1 f(x_1, x_2) dx_1 dx_2 \\ &= 0.064\end{aligned}$$

- To compute probability that $X_1 \geq 1$ and $X_2 \geq 2$, we construct event $\mathcal{A} = \{1 \leq x_1 \leq \infty \ \& \ 2 \leq x_2 \leq \infty\}$ and we have that:

$$\begin{aligned}\mathbb{P}(X \in \mathcal{A}) &= \int_1^\infty \int_2^\infty f(x_1, x_2) dx_1 dx_2 \\ &= 0.30.\end{aligned}$$

Example: Flow Cytometry `ch3_flow_cytometer_discrete.m`

- A microemulsion with 10,000 droplets is passed through a flow cytometer.
- Cytometer reports side scatter (SSC) and front scatter (FSC) values for droplets.
- For simplicity, the SSC and FSC values are assumed to belong to three possible categories - low (L), medium (M), and high (H). In other words, the FSC and SSC are treated as discrete RVs.
- Use data to obtain the joint pdf for the FSC and SSC.
- Determine what is the most likely category that the droplets belong to.



Example: Flow Cytometry `ch3_flow_cytometer_discrete.m`

- Define RV $X = (X_1, X_2)$ with X_1=FSC and X_2=SSC; the individual domains of the RVs are $\mathcal{D}_1 = \{L, M, H\}$ and $\mathcal{D}_2 = \{L, M, H\}$, and the joint domain is:

$$\begin{aligned}\mathcal{D} &= \mathcal{D}_1 \times \mathcal{D}_2 \\ &= \{(L,L),(L,M),(L,H),(M,L),(M,M),(M,H),(H,L),(H,M),(H,H)\}.\end{aligned}$$

- To obtain joint pdf, we count number of droplets (observations) in each value of the domain and normalize them by the total number of observations $S = 10,000$.

- The (empirical) joint pdf can expressed as a matrix of the form:

$\hat{f}(x_1, x_2) =$

	L	M	H
H	0.0028	0.0239	0.0012
M	0.5994	0.1114	0
L	0.2613	0	0

Here, the rows correspond to SSC values and columns to the FSC values.

- One can confirm that all the entries in the matrix add up to one.

- Most likely category is (L, M), which as a joint pdf $\hat{\mathbb{P}}(X_1 = L, X_2 = M) = 0.5994$. This means that 5,994 droplets lie on this category.

Joint PDFs and CDFs

- *Joint cdf* for $X = (X_1, X_2)$ is:

$$F(x_1, x_2) = \mathbb{P}(X_1 \leq x_1 \,\&\, X_2 \leq x_2), \quad (x_1, x_2) \in \mathcal{D}$$

- For continuous X:

$$F(t_1, t_2) = \int_{(x_1, x_2) \in \mathcal{F}} f(x_1, x_2) dx_1 dx_2$$

with domain $\mathcal{F} = \mathcal{D} \cap \{x_1 \leq t_1 \,\&\, x_2 \leq t_2\}$ and:

$$f(x_1, x_2) = \frac{\partial}{\partial x_2}\left(\frac{\partial F(x_1, x_2)}{\partial x_1}\right) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}, \quad (x_1, x_2) \in \mathcal{D}.$$

- Interpretation of derivative as limit of an infinitesimal small domain is:

$$f(x_1, x_2) = \lim_{\Delta x_2 \to 0} \frac{1}{\Delta x_2}\left(\frac{dF(x_1, x_2 + \Delta x_2)}{dx_1} - \frac{dF(x_1, x_2)}{dx_1}\right)$$

where

$$\frac{dF(x_1, x_2 + \Delta x_2)}{dx_1} = \lim_{\Delta x_1 \to 0} \frac{F(x_1 + \Delta x_1, x_2 + \Delta x_2) - F(x_1, x_2 + \Delta x_2)}{\Delta x_1}$$

$$\frac{dF(x_1, x_2)}{dx_1} = \lim_{\Delta x_1 \to 0} \frac{F(x_1 + \Delta x_1, x_2) - F(x_1, x_2)}{\Delta x_1}.$$

Joint PDFs and CDFs

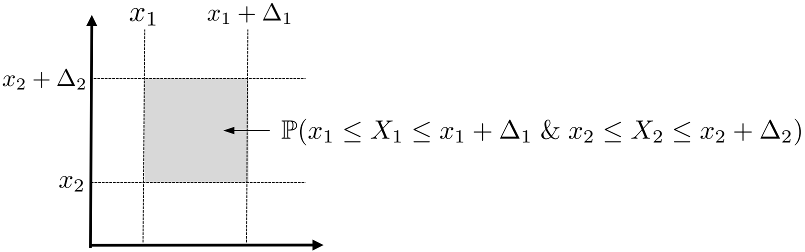
- This suggests discrete approximation:

$$f(x_1, x_2) \approx \frac{1}{\Delta x_1 \Delta x_2} \left[\; F(x_1 + \Delta x_1, x_2 + \Delta x_2) - F(x_1, x_2 + \Delta x_2) \right. \\ \left. F(x_1, x_2) - F(x_1 + \Delta x_1, x_2) \right].$$

- One can show that this reduces to:

$$f(x_1, x_2)\Delta x_1 \Delta x_2 \approx \mathbb{P}(x_1 \leq X_1 \leq x_1 + \Delta x_1 \;\&\; x_2 \leq X_2 \leq x_2 + \Delta x_2).$$

- Joint pdf thus gives prob that X is in box $[x_1, x_1 + \Delta x_1] \times [x_2, x_2 + \Delta x_2]$.



Example: Lifetime of Thermostat

Determine cdf for thermostat and prob that lifetime of sensor *and* processor is $\leq 1yr$.

- Domain is $\mathcal{D} = \{0 \leq x_1 \leq \infty \;\&\; 0 \leq x_2 \leq \infty\}$.
- We have $\mathcal{F} = \mathcal{D} \cap \{x_1 \leq t_1 \,\&\, x_2 \leq t_2\}$ and thus $\mathcal{F} = \{0 \leq x_1 \leq t_1 \;\&\; 0 \leq x_2 \leq t_2\}$.
- Cdf is:

$$\begin{aligned} F(t_1, t_2) &= \int_0^{t_1} \int_0^{t_2} f(x_1, x_2) dx_1 dx_2 \\ &= \left(e^{-0/\mu_1} - e^{-t_1/\mu_1}\right)\left(e^{-0/\mu_2} - e^{-t_2/\mu_2}\right) \\ &= \left(1 - e^{-t_1/\mu_1}\right)\left(1 - e^{-t_2/\mu_2}\right). \end{aligned}$$

- We use this to compute $\mathbb{P}(X_1 \leq t_1 \;\&\; X_2 \leq t_2) = F(t_1, t_2) = 0.06$.
- Prob that both components fail before one year is rather small.

Joint PDFs and CDFs

- As in univariate case, we can construct *empirical* approximations for pdf and cdf.
- Given data $x_\omega,\ \omega \in \mathcal{S}$, empirical approx of cdf is:

$$\begin{aligned}\hat{F}(t_1, t_2) &= \hat{\mathbb{P}}(X_1 \le t_1 \,\&\, X_2 \le t_2)\\ &= \frac{1}{S}\sum_{\omega\in\mathcal{S}} \mathbf{1}[x_{\omega,1} \le t_1 \,\&\, x_{\omega,2} \le t_2], \quad (t_1, t_2) \in \hat{\mathcal{D}}.\end{aligned}$$

- Indicator is equal to one if observation x_ω satisfies $x_{\omega,1} \le t_1$ and $x_{\omega,2} \le t_2$.
- Can obtain empirical pdf by using a approximation of cdf:

$$\hat{f}(x_1, x_2) \approx \frac{1}{\Delta x_1 \Delta x_2}\hat{\mathbb{P}}(x_1 \le X_1 \le x_1 + \Delta x_1 \,\&\, x_2 \le X_2 \le x_2 + \Delta x_2)$$

Example: Flow Cytometry `ch3_flow_cytometer_empiricial.m`

- Treating the FSC and SSC as discrete facilitates analysis but more appropriate to treat these as continuous.
- Use flow cytometer data to obtain an empirical joint pdf of FSC and SSC.
- To construct the approximation, we first determine the empirical domain covered by the observations. From the data, we find:

$$\hat{\mathcal{D}} = \{4.2365 \leq X_1 \leq 6.0834 \;\&\; 3.6191 \leq X_2 \leq 5.8543\}.$$

- We partition (discretize) box by using bins of size $\Delta x_1 = \Delta x_2 = 0.05$.
- This gives 46 bins in SSC domain and 38 bins in FSC domain (total of 1748 bins).
- To construct empirical pdf, we count number of observations in each of bins.

Example: Flow Cytometry `ch3_flow_cytometer_empiricial.m`

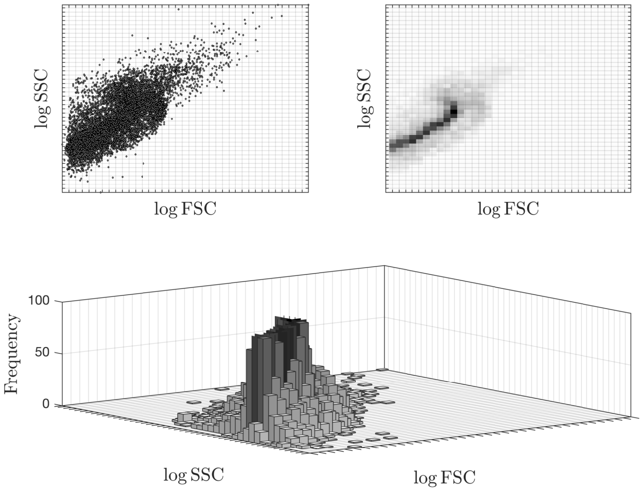


Figure: Top-left: Scatter plot showing observations. Top-right: Density of observations in bins (darker is denser). Bottom: Empirical pdf shown as 2D histogram.

Marginal PDFs

- Want probs for X_1 *regardless* of what value X_2 takes (and other way around).
- Probs are obtained from *marginal pdfs*. For continuous RV:

$$f_1(x_1) = \int_{x_2 \in \mathcal{D}_2} f(x_1, x_2)dx_2,\ x_1 \in \mathcal{D}_1$$
$$f_2(x_2) = \int_{x_1 \in \mathcal{D}_1} f(x_1, x_2)dx_1,\ x_2 \in \mathcal{D}_2$$

- Marginal pdfs are obtained by "integrating out" effect of the RV that we ignore.
- Consider we want $\mathbb{P}(X_1 \in \mathcal{A}_1)$ regardless of what value X_2 takes. This is:

$$\mathbb{P}(X_1 \in \mathcal{A}_1) = \int_{x_1 \in \mathcal{A}_1} \overbrace{\int_{x_2 \in \mathcal{D}_2} f(x_1, x_2)dx_2}^{f_1(x_1)} dx_1$$
$$= \int_{x_1 \in \mathcal{A}_1} f_1(x_1)dx_1$$

- Marginal pdfs $f_1(x_1)$ and $f_2(x_2)$ have associated marginal cdfs $F_1(x_1), F_2(x_2)$.

Example: Lifetime of Thermostat

- Marginal pdf and cdf of component lifetimes X_1, X_2.
- Compute prob that X_1 survives after 1 year.
- Compute prob that X_2 survives after 1 year.

- Marginal pdf for X_1 is obtained by integrating X_2 out over $\mathcal{D}_2 = [0, \infty)$:

$$f_1(x_1) = \int_0^\infty f(x_1, x_2) dx_2 = \int_0^\infty \left(\frac{1}{\mu_1 \cdot \mu_2} e^{-(x_1/\mu_1 + x_2/\mu_2)} \right) dx_2$$
$$= \frac{1}{\mu_1 \cdot \mu_2} e^{-x_1/\mu_1} \left(\int_0^\infty e^{-x_2/\mu_2} dx_2 \right) = \frac{1}{\mu_1} e^{-x_1/\mu_1}.$$

- Applying a similar procedure we obtain:

$$f_2(x_2) = \frac{1}{\mu_2} e^{-x_2/\mu_2}.$$

Example: Lifetime of Thermostat

- Marginals are pdfs of exponential RVs with $\mu_1 = 1$ and $\mu_2 = 10$ years and thus:

$$F_1(x_1) = 1 - e^{-x_1/\mu_1} \qquad F_2(x_2) = 1 - e^{-x_2/\mu_2}.$$

- Probabilities that components survive after one year are:

$$\mathbb{P}(X_1 \geq 1) = 1 - F_1(1) = e^{-1/1} = 0.37$$
$$\mathbb{P}(X_2 \geq 1) = 1 - F_2(1) = e^{-1/10} = 0.90.$$

Conditional PDFs and CDFs

- Want probs for X_1 *given knowledge* that X_2 takes value x_2 (or other way around).

- These probabilities are obtained from *conditional pdfs*:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}, \; x_1 \in \mathcal{D}_1$$
$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}, \; x_2 \in \mathcal{D}_2$$

  i.e., $f(x_1|x_2) = \mathbb{P}(X_1 = x_1|X_2 = x_2)$ and $f(x_2|x_1) = \mathbb{P}(X_2 = x_2|X_1 = x_1)$.

- Note resemblance of conditional pdfs with conditional probabilities seen previously.

Conditional PDFs and CDFs

- These expressions can also be written as:

$$f_2(x_2)f(x_1|x_2) = f(x_1, x_2),\ x_1 \in \mathcal{D}_1$$
$$f_1(x_1)f(x_2|x_1) = f(x_1, x_2),\ x_2 \in \mathcal{D}_2$$

- Consider we want $\mathbb{P}(a_1 \leq X_1 \leq b_1 | X_2 = x_2)$ this is given by:

$$\mathbb{P}(a_1 \leq X_1 \leq b_1 | X_2 = x_2) = \int_{a_1}^{b_1} f(x_1|x_2)dx_1$$

- Joint pdfs have associated marginal cdfs $F(x_1|x_2)$ and $F(x_2|x_1)$.

Example: Lifetime of Thermostat

- Obtain the conditional pdf and cdf of component lifetimes X_1, X_2.
- Compute prob that X_1 survives after 1 year given that X_2 survives 1 year.
- Compute prob that X_2 survives after 1 year given that X_1 survives 1 year.
- Compare probs against the marginal probabilities of survival.

Example: Lifetime of Thermostat

- Conditional pdfs are obtained from joint and marginal pdfs as:

$$f(x_1|x_2) = \frac{f(x_1,x_2)}{f_2(x_2)} = \frac{\frac{1}{\mu_1\cdot\mu_2}e^{-(x_1/\mu_1+x_2/\mu_2)}}{\frac{1}{\mu_2}e^{-x_2/\mu_2}}$$
$$= \frac{1}{\mu_1}e^{-x_1/\mu_1}.$$

- We apply a similar procedure to obtain:

$$f(x_2|x_1) = \frac{1}{\mu_2}e^{-x_2/\mu_2}.$$

- Note $f(x_1|x_2)$ does not depend on x_2 and $f(x_2|x_1)$ does not depend on x_1.
- Knowing X_2 does not influence probability of failure of X_1 and viceversa.
- Note conditional pdfs are marginal pdfs of X_1 and X_2.
- Thus have $\mathbb{P}(X_1 \geq 1|X_2 = 1) = \mathbb{P}(X_1 \geq 1) = 0.37$ and $\mathbb{P}(X_2 \geq 1|X_1 = 1) = \mathbb{P}(X_2 \geq 1) = 0.90$.

Independence

Conditional pdfs tell us how knowledge in one RV resolves uncertainty in another (i.e., how much knowledge of one is embedded in the other).

So what if knowledge of one does not resolve uncertainty of the other?
This gives rise to concept of *independence*.

- RVs X_1 and X_2 are said to be independent if:

$$f(x_1|x_2) = f_1(x_1),\ x_1 \in \mathcal{D}_1$$
$$f(x_2|x_1) = f_2(x_2),\ x_2 \in \mathcal{D}_2$$

- This implies that:

$$f(x_1, x_2) = f_2(x_2) f_1(x_1), \quad (x_1, x_2) \in \mathcal{D}$$

- For discrete X this is equivalent to:

$$\mathbb{P}(X_1 = x_1 \,\&\, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)$$

- For continuous X this is interpreted in infinitesimal sense.

Example: Lifetime of Thermostat

Consider the thermostat example and address the following:

- Prove that lifetimes X_1 and X_2 are independent and provide a physical explanation.
- Prove that joint cdf is the product of marginal cdfs.

Example: Lifetime of Thermostat

- Condition for independence is $f(x_1, x_2) = f_1(x_1) f_2(x_2)$.
- From previous examples we have that:

$$f(x_1, x_2) = \frac{1}{\mu_1 \cdot \mu_2} e^{-(x_1/\mu_1 + x_2/\mu_2)} = \left(\frac{1}{\mu_1} e^{-x_1/\mu_1} \right) \left(\frac{1}{\mu_2} e^{-x_2/\mu_2} \right),$$

  and,

$$f_1(x_1) f_2(x_2) = \left(\frac{1}{\mu_1} e^{-x_1/\mu_1} \right) \left(\frac{1}{\mu_2} e^{-x_2/\mu_2} \right).$$

- Failure of components might be independent if they do not interact physically.
- Joint cdf is $F(x_1, x_2) = (1 - e^{-x_1/\mu_1})(1 - e^{-x_2/\mu_2})$ and marginals are $F_1(x_1) = 1 - e^{-x_1/\mu_1}$ and $F_2(x_2) = 1 - e^{-x_2/\mu_2}$.
- We thus have $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$ and this implies $\mathbb{P}(X_1 \leq x_1 \ \& \ X_2 \leq x_2) = \mathbb{P}(X_1 \leq x_1) \cdot \mathbb{P}(X_2 \leq x_2)$.

Conditional Events

- Focused on probs that are conditional to simple events (e.g., $X_1 = x_1$, $X_2 \leq x_2$).
- Probs that are conditional of general events can be derived from domain properties.
- Return to definition of conditional probability:

$$\mathbb{P}(X \in \mathcal{A}_1 | X \in \mathcal{A}_2) = \frac{\mathbb{P}(X \in \mathcal{A}_1 \ \& \ X \in \mathcal{A}_2)}{\mathbb{P}(X \in \mathcal{A}_2)}.$$

  where $\mathcal{A}_1$ and $\mathcal{A}_2$ are subdomains of $\mathcal{D}$ that capture specific events.
- $\mathbb{P}(X \in \mathcal{A}_1 | X \in \mathcal{A}_2)$ is prob that event $\mathcal{A}_1$ occurs given that $\mathcal{A}_2$ has occurred.
- $\mathbb{P}(X \in \mathcal{A}_1 \ \& \ X \in \mathcal{A}_2)$ is joint prob of $\mathcal{A}_1$ and $\mathcal{A}_2$, and $\mathbb{P}(X \in \mathcal{A}_2)$ is marginal prob.

Conditional Events

- For continuous X, prob $\mathbb{P}(X \in \mathcal{A}_1 | X \in \mathcal{A}_2)$ is:

$$\mathbb{P}(X \in \mathcal{A}_1 | X \in \mathcal{A}_2) = \frac{\displaystyle\int_{x \in \mathcal{A}_1 \cap \mathcal{A}_2} f(x)dx}{\displaystyle\int_{x \in \mathcal{A}_2} f(x)dx}.$$

- Recall that joint event $\{X \in \mathcal{A}_1 \ \& \ X \in \mathcal{A}_2\}$ is equivalent to $X \in \mathcal{A}_1 \cap \mathcal{A}_2$.

- Conditional prob $\mathbb{P}(X_1 = x_1 | X_2 = x_2)$ is a special case with events of the form $\mathcal{A}_1 = \{X_1 = x_1 \ \& \ X_2 \in \mathcal{D}_2\}$ and $\mathcal{A}_2 = \{X_1 \in \mathcal{D}_1 \ \& \ X_2 = x_2\}$.

Conditional Events

- Recall that, if $\mathcal{A}_1$ and $\mathcal{A}_2$ are *independent* events, then:

$$\mathbb{P}(X \in \mathcal{A}_1 \;\&\; X \in \mathcal{A}_2) = \mathbb{P}(X \in \mathcal{A}_1) \cdot \mathbb{P}(X \in \mathcal{A}_2).$$

- This implies that:

$$\int_{x \in \mathcal{A}_1 \cap \mathcal{A}_2} f(x)dx = \left(\int_{x \in \mathcal{A}_1} f(x)dx \right) \cdot \left(\int_{x \in \mathcal{A}_2} f(x)dx \right)$$

- This explains why joint cdf is product of marginal cdfs for X_1, X_2 independent.

Conditional Events

- Prob that event $\mathcal{A}_1$ *or* $\mathcal{A}_2$ occur is given by:

$$\mathbb{P}(X \in \mathcal{A}_1 \cup \mathcal{A}_2) = \mathbb{P}(X \in \mathcal{A}_1) + \mathbb{P}(X \in \mathcal{A}_2) - \mathbb{P}(X \in \mathcal{A}_1 \cap \mathcal{A}_2).$$

- This is computed by using joint and marginal pdfs of X as:

$$\mathbb{P}(X \in \mathcal{A}_1 \cup \mathcal{A}_2) = \int_{x \in \mathcal{A}_1} f(x)dx + \int_{x \in \mathcal{A}_2} f(x)dx - \int_{x \in \mathcal{A}_1 \cap \mathcal{A}_2} f(x)dx.$$

- Pdfs and event properties can be used to compute probs of complex events.

Example: Lifetime of Thermostat

Compute probability that sensor or processor last more than a year.

- We build events $\mathcal{A}_1 = \{x_1 \geq 1 \ \& \ x_2 \in \mathcal{D}_2\}$ and $\mathcal{A}_2 = \{x_1 \in \mathcal{D}_1 \ \& \ x_2 \geq 1\}$, which give $\mathcal{A}_1 \cap \mathcal{A}_2 = \{x_1 \geq 1 \ \& \ x_2 \geq 1\}$, and $\mathcal{A}_1 \cup \mathcal{A}_2 = \{x_1 \geq 1 \text{ or } x_2 \geq 1\}$.

- Computing probabilities we have:

$$\begin{aligned} \int_{x \in \mathcal{A}_1} f(x) dx &= \int_1^\infty \int_0^\infty f(x_1, x_2) dx_1 dx_2 = 0.37 \\ \int_{x \in \mathcal{A}_2} f(x) dx &= \int_0^\infty \int_1^\infty f(x_1, x_2) dx_1 dx_2 = 0.90 \\ \int_{x \in \mathcal{A}_1 \cap \mathcal{A}_2} f(x) dx &= \int_1^\infty \int_1^\infty f(x_1, x_2) dx_1 dx_2 = 0.33. \end{aligned}$$

  and thus $\mathbb{P}(X \in \mathcal{A}_1 \cup \mathcal{A}_2) = 0.37 + 0.90 - 0.33 = 0.94$.

- There is a high probability that either one of them lasts more than a year.

Summarizing Statistics

As with univariate RVs, we are often interested in characterizing multivariate RVs by using summarizing statistics. A few new concepts will emerge in this discussion.

- For continuous RV, *joint expectation* of $X = (X_1, X_2)$ is vector $\mathbb{E}[X] \in \mathbb{R}^2$:

$$\mathbb{E}[X] = \mathbb{E}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}.$$

- This vector is computed as:

$$\mathbb{E}[X] = \int_{x \in \mathcal{D}} x f(x) dx.$$

- If we write $x = (x_1, x_2)$, we can express expected value as:

$$\mathbb{E}[X] = \int_{(x_1, x_2) \in \mathcal{D}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} f(x_1, x_2) dx_1 dx_2.$$

- We can thus write $\mathbb{E}[X] = (\mathbb{E}[X_1], \mathbb{E}[X_2])$ with:

$$\mathbb{E}[X_1] = \int_{(x_1, x_2) \in \mathcal{D}} x_1 f(x_1, x_2) dx_1 dx_2$$

$$\mathbb{E}[X_2] = \int_{(x_1, x_2) \in \mathcal{D}} x_2 f(x_1, x_2) dx_1 dx_2.$$

Summarizing Statistics

- We highlight difference between joint expectations $\mathbb{E}[X_1]$ and $\mathbb{E}[X_2]$ and marginal expectations of X_1 and X_2.

- The *marginal expectations* are given by:

$$\mathbb{E}_m[X_1] = \int_{x_1 \in \mathcal{D}_1} x_1 f_1(x_1) dx_1$$
$$\mathbb{E}_m[X_2] = \int_{x_2 \in \mathcal{D}_2} x_2 f_2(x_2) dx_2.$$

  In some special situations, joint and marginal expectations are the same (e.g., X_1 and X_2 are independent).

- Joint expectation of scalar function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ of X is a scalar value:

$$\mathbb{E}[\varphi(X)] = \int_{(x_1,x_2)\in\mathcal{D}} \varphi(x_1, x_2) f(x_1, x_2) dx_1 dx_2$$

- *Conditional* expectation of X_1 (given $X_2 = x_2$) is:

$$\mathbb{E}[X_1|X_2 = x_2] = \int_{x_1 \in \mathcal{D}_1} x_1 f(x_1|x_2) dx_1.$$

- Joint expectation of n-dim $X = (X_1, X_2, ..., X_n)$ is vector $\mathbb{E}[X] \in \mathbb{R}^n$:

$$\mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix}.$$

where:

$$\mathbb{E}[X_i] = \int_{x \in \mathcal{D}} x_i f(x) dx,\ i = 1, ..., n$$

- It is possible to compute expectations that are conditional on events. For example:

$$\mathbb{E}[\varphi(X) | X \in \mathcal{A}] = \int_{x \in \mathcal{A}} \varphi(x) f(x) dx,$$

Summarizing Statistics

For univ X, we seek to understand variability of RV because this is a measure of its uncertainty.

For multiv X, we seek to understand variability of X_1 and X_2 and to understand how X_1 and X_2 vary together.

- *Covariance* of $X = (X_1, X_2)$ is matrix $\mathrm{Cov}(X) \in \mathbb{R}^{2\times 2}$:

$$\mathrm{Cov}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right]$$

  It is important to highlight that here we use the *joint expectation*.

- To simplify, we define $m = \mathbb{E}[X]$ with entries (m_1, m_2) and thus:

$$\begin{aligned}
\mathrm{Cov}[X] &= \mathbb{E}\left[(X - m)(X - m)^T\right] \\
&= \mathbb{E}\left[\begin{bmatrix} X_1 - m_1 \\ X_2 - m_2 \end{bmatrix} [X_1 - m_1 \; X_2 - m_2]\right] \\
&= \begin{bmatrix} \mathbb{E}[(X_1 - m_1)^2] & \mathbb{E}[(X_1 - m_1)(X_2 - m_2)] \\ \mathbb{E}[(X_2 - m_2)(X_1 - m_1)] & \mathbb{E}[(X_2 - m_2)^2] \end{bmatrix}
\end{aligned}$$

Summarizing Statistics

- We define the individual entries of the matrix as follows:

$$\begin{aligned}
\mathrm{Cov}[X]_{11} &= \mathbb{E}[(X_1 - m_1)^2] \\
\mathrm{Cov}[X]_{12} &= \mathbb{E}[(X_1 - m_1)(X_2 - m_2)] \\
\mathrm{Cov}[X]_{21} &= \mathbb{E}[(X_2 - m_2)(X_1 - m_1)] \\
\mathrm{Cov}[X]_{22} &= \mathbb{E}[(X_2 - m_2)^2].
\end{aligned}$$

- Note $\mathrm{Cov}[X]_{12} = \mathrm{Cov}[X]_{21}$ (covariance matrix is *symmetric*) and

$$\begin{aligned}
\mathrm{Cov}[X]_{11} &= \mathbb{V}[X_1] \\
\mathrm{Cov}[X]_{22} &= \mathbb{V}[X_2].
\end{aligned}$$

- The covariance matrix is also positive semi-definite (all its eigenvalues are non-negative).

- Variance $\mathrm{Cov}[X]_{11}$ tells us how, on average, X_1 varies around m_1; $\mathrm{Cov}[X]_{22}$ tells us how, on average, X_2 varies around m_2.

Summarizing Statistics

- Covariance $\mathrm{Cov}[X]_{12} = \mathrm{Cov}[X]_{21}$ tells us how, *on average*, X_1 varies with X_2:
  - If $\mathrm{Cov}[X]_{12} > 0$ indicates that X_1 and X_2 move in the same direction (on average).
  - If $\mathrm{Cov}[X]_{12} < 0$ indicates that X_1 and X_2 move in opposite directions (on average).
  - If $\mathrm{Cov}[X]_{12} = 0$ indicates that X_1 and X_2 do not move together (on average). Does this imply independence?
- *Covariance* $\mathrm{Cov}[X]_{12}$ is measure of *interdependence* between RVs (connectivity).
- For instance, if $X_2 = \alpha X_1$, then $\mathrm{Cov}[X]_{12} = \alpha \mathbb{V}[X_1]$.
- **Important:** Physical connections between RVs are encoded in covariance.

Summarizing Statistics

- *Autocorrelation* matrix is a summarizing statistic used to measure variability:
$$\mathrm{ACorr}[X] = \mathbb{E}[XX^T] = \begin{bmatrix} \mathbb{E}[X_1X_1] & \mathbb{E}[X_1X_2] \\ \mathbb{E}[X_2X_1] & \mathbb{E}[X_2X_2] \end{bmatrix}$$
Definition similar to covariance but do not use expected value as reference.

- Autocorrelation matrix is related to covariance matrix as:
$$\mathrm{Cov}(X) = \mathrm{ACorr}[X] - \mathbb{E}[X]\mathbb{E}[X]^T.$$

- *Pearson* correlation matrix is defined as:
$$\mathrm{PCorr}[X] = D^{-1}\mathrm{Cov}[X]D^{-1}$$
where $D = \sqrt{\mathrm{diag}(\mathrm{Cov}[X])}$.

- Can show that entries of Pearson matrix satisfy:
$$\begin{aligned} \mathrm{PCorr}[X]_{11} &= 1 \\ \mathrm{PCorr}[X]_{12} &\in [-1, 1] \\ \mathrm{PCorr}[X]_{21} &\in [-1, 1] \\ \mathrm{PCorr}[X]_{22} &= 1. \end{aligned}$$

Summarizing Statistics

- Generalizing covariance and correlation to n-dimensional case is straightforward.
- Covariance of $X = (X_1, X_2, ..., X_n)$ is:

$$\mathrm{Cov}[X] = \begin{bmatrix} \mathrm{Cov}[X]_{11} & \mathrm{Cov}[X]_{12} & \cdots & \mathrm{Cov}[X]_{1n} \\ \mathrm{Cov}[X]_{21} & \mathrm{Cov}[X]_{22} & \cdots & \mathrm{Cov}[X]_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[X]_{n1} & \mathrm{Cov}[X]_{n2} & \cdots & \mathrm{Cov}[X]_{nn} \end{bmatrix}$$

where:

$$\mathrm{Cov}[X]_{ij} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])],\ i, j = 1, ..., n.$$

- Peason correlation of $X = (X_1, X_2, ..., X_n)$ is:

$$\mathrm{PCorr}[X] = \begin{bmatrix} 1 & \mathrm{PCorr}[X]_{12} & \cdots & \mathrm{PCorr}[X]_{1n} \\ \mathrm{PCorr}[X]_{21} & 1 & \cdots & \mathrm{PCorr}[X]_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Pcorr}[X]_{n1} & \mathrm{PCorr}[X]_{n2} & \cdots & 1 \end{bmatrix}$$

where:

$$\mathrm{PCorr}[X]_{ij} = \frac{\mathrm{Cov}[X]_{ij}}{\mathbb{SD}[X_i]\mathbb{SD}[X_j]},\ i = 1, ..., n,\ j = 1, ..., n.$$

Summarizing Statistics

- We can use data to compute empirical estimates for covariance and correlation matrices.
- Imagine we have observations $x_\omega = (x_{\omega,1}, x_{\omega,2}, ..., x_{\omega,n})$, $\omega \in \mathcal{S}$ of X.
- Empirical covariance matrix (also known as the sample covariance matrix) is:

$$\hat{\mathrm{Cov}}[X]_{ij} = \frac{1}{S} \sum_{\omega \in \mathcal{S}} (x_{\omega,i} - \hat{m}_i)(x_{\omega,j} - \hat{m}_j),\ i, j = 1, ..., n$$

where

$$\hat{m}_i = \hat{\mathbb{E}}[X_i] = \frac{1}{S} \sum_{\omega \in \mathcal{S}} x_{\omega,i},\ i = 1, ..., n.$$

- Here, $\hat{m}_i$ are empirical expectations of X_i, $i = 1, ..., n$.

Example: Covariance and Correlation for Gibbs Reactor `ch3_gibbs_covariance.m`

- Revisit reactor under which the reaction $CO + 2H_2 \leftrightarrow CH_3OH$ takes place.
- Recall that this reaction is favored at high pressure.
- Have data available for $n = 5$ RVs: pressure, conversion, output flow for reactants (CO, H_2) and for product CH_3OH.

Consider the following questions:

- Do you expect positive or negative corr between conversion and pressure?
- Do you expect positive or negative corr between pressure and output flow of CO?
- How are the output flows for CO, H_2, and CH_3OH related?

Example: Covariance and Correlation for Gibbs Reactor `ch3_gibbs_covariance.m`

- Without even looking at the data, our physical understanding of reactor gives us so intuition as to how the variables are connected between each other.

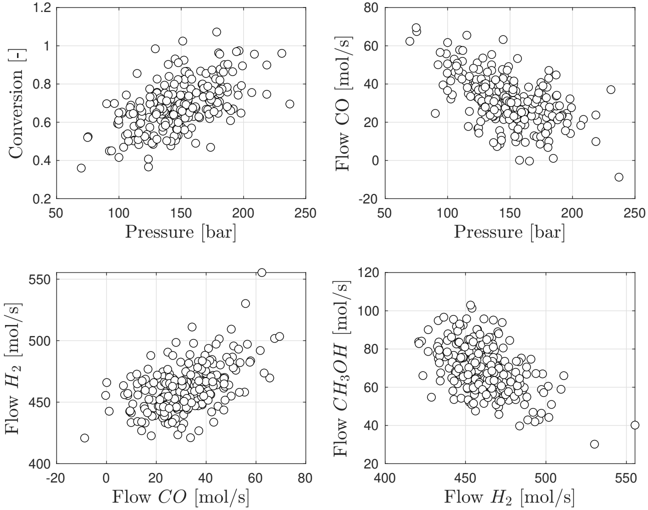
- We know that reaction is favored by high pressure and thus expect that conversion is positively correlated with pressure.

- From conservation laws, we also know that output flows of reactants decrease as we increase the output flow of product (with higher conversion).

- We thus expect that output flow of CO is negatively correlated with pressure (knowledge of one variable is encoded in another variable).

Example: Covariance and Correlation for Gibbs Reactor `ch3_gibbs_covariance.m`

Interdependencies are confirmed visually from data but trends hidden by noise.



Example: Covariance and Correlation for Gibbs Reactor `ch3_gibbs_covariance.m`

- Use data for computing sample covariance and Pearson correlation matrices.

- Covariance matrix is:

$$\hat{\mathrm{Cov}}[X] = \begin{bmatrix} 776.94 & 1.83 & -195.91 & -396.57 & 181.45 \\ 1.83 & 0.02 & -0.54 & -1.11 & 0.40 \\ -195.91 & -0.54 & 163.81 & 117.59 & -58.17 \\ -396.57 & -1.11 & 117.59 & 338.35 & -116.12 \\ 181.45 & 0.40 & -58.17 & -116.12 & 161.70 \end{bmatrix}.$$

- Rows and columns are arranged as: 1) pressure, 2) conversion, and 3,4,5) output flows of CO, H_2, and CH_3OH.

- Sample correlation matrix is:

$$\mathrm{P}\hat{\mathrm{Corr}}[X] = \begin{bmatrix} 1.00 & 0.53 & -0.55 & -0.77 & 0.51 \\ 0.53 & 1.00 & -0.34 & -0.49 & 0.25 \\ -0.55 & -0.34 & 1.00 & 0.50 & -0.36 \\ -0.77 & -0.49 & 0.50 & 1.00 & -0.50 \\ 0.51 & 0.25 & -0.36 & -0.50 & 1.00 \end{bmatrix}$$

- Sample covariance and correlation matrix confirm physical intuition.

Multivariate Gaussian Model

- By far, the most used multivariate RV model is the *Gaussian* RV model.

- Joint pdf of an n-dimensional Gaussian $X \sim \mathcal{N}(\mu, \Sigma)$ is:

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right), \ x \in \mathcal{D},$$

- Parameters of this model are $\mu \in \mathbb{R}^n$ (a vector) and $\Sigma \in \mathbb{R}^{n\times n}$ (a matrix).

- *Domain* of Gaussian X is $\mathcal{D} = (-\infty, \infty)^n \subseteq \mathbb{R}^n$.

- Gaussian RV has key property that $\mu = \mathbb{E}[X]$ and $\Sigma = \mathrm{Cov}[X]$.

Multivariate Gaussian Model

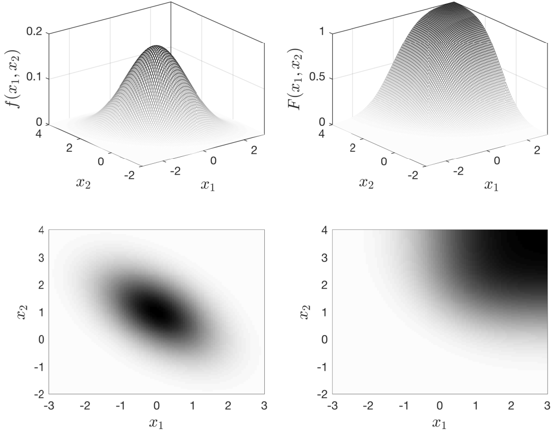


Figure: Joint pdf and joint cdf for multivariate Gaussian and 2D projections.

- Pdf becomes concentrated (dense) in center (indicating a higher pdf value).
- Density forms ellipsoidal regions (diffuses as moving from center).

Multivariate Gaussian Model

- For $n = 2$, joint pdf can be written as:

$$f(x_1, x_2) = c \cdot \exp\left(-\frac{1}{2}\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\right)$$

where:

$$c = \frac{1}{2\pi \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix}^{1/2}}.$$

- Recall $\mathbb{E}[X] = \mu$ and $\mathrm{Cov}[X] = \Sigma$.
- If (X_1, X_2) are independent, we have $\Sigma_{12} = \Sigma_{21} = 0$ and thus:

$$f(x_1, x_2) = \frac{1}{\sqrt{2\pi\Sigma_{11}}} \exp\left(\frac{-(x_1 - \mu_1)^2}{2\Sigma_{11}}\right) \cdot \frac{1}{\sqrt{2\pi\Sigma_{22}}} \exp\left(\frac{-(x_2 - \mu_2)^2}{2\Sigma_{22}}\right)$$

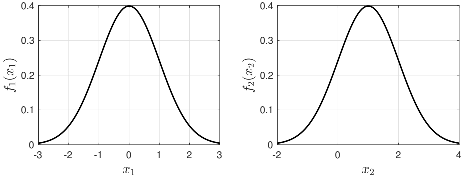
Multivariate Gaussian Model

- Marginals of Gaussian $X = (X_1, X_2)$ are:

$$f_1(x_1) = \frac{1}{\sqrt{2\pi\Sigma_{11}}} \exp\left(\frac{-(x_1 - \mu_1)^2}{2\Sigma_{11}}\right)$$
$$f_2(x_2) = \frac{1}{\sqrt{2\pi\Sigma_{22}}} \exp\left(\frac{-(x_2 - \mu_2)^2}{2\Sigma_{22}}\right).$$

- These correspond to pdfs of $X_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$ and $X_2 \sim \mathcal{N}(\mu_2, \Sigma_{22})$.

- Here are marginal pdfs for Gaussian with $\mu = (0, 1)$ and $\Sigma = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$.



- These correspond to pdfs of $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(1, 1)$.

Multivariate Gaussian Model

- Conditional pdfs of Gaussian $X = (X_1, X_2)$ are:

$$f(x_1|x_2) = \frac{1}{\sqrt{2\pi\Sigma_{1|2}}} \exp\left(\frac{-(x_1 - \mu_{1|2})^2}{2\Sigma_{1|2}}\right)$$
$$f(x_2|x_1) = \frac{1}{\sqrt{2\pi\Sigma_{2|1}}} \exp\left(\frac{-(x_2 - \mu_{2|1})^2}{2\Sigma_{2|1}}\right).$$

where:

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$
$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$
$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)$$
$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

- We thus have $X_1|X_2 \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$ and $X_2|X_1 \sim \mathcal{N}(\mu_{2|1}, \Sigma_{2|1})$.

- This provides important insights on how covariance between X_1 and X_2 affects to what degree knowledge of one resolves uncertainty of the other.

Example:Uncertainty Reduction in Gibbs Reactor (Part I).

Pressure and conversion (X_1, X_2) are Gaussian with covariance and correlation:

$$\mathrm{Cov}[X] = \begin{bmatrix} 776.94 & 1.83 \\ 1.83 & 0.02 \end{bmatrix}$$

$$\mathrm{PCorr}[X] = \begin{bmatrix} 1 & 0.53 \\ 0.53 & 1 \end{bmatrix}.$$

- How does having knowledge in pressure resolves uncertainty in conversion?
- How does having knowledge in conversion resolves uncertainty in pressure?

Example:Uncertainty Reduction in Gibbs Reactor.

- To quantify by how much is uncertainty in conversion reduced by having knowledge of pressure, we compute variance of conditional pdf $f(x_2|x_1)$.

- Variance of $X_2|X_1$ is given by:

$$\begin{aligned}\Sigma_{2|1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\ &= 0.02 - 0.53 \cdot (776.94)^{-1} \cdot 0.53 \\ &= 0.019\end{aligned}$$

  This is a relative reduction of *5%*.

- To quantify how knowledge of conversion affects uncertainty in pressure, we obtain variance of $X_1|X_2$:

$$\begin{aligned}\Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= 776.94 - 1.83 \cdot (0.02)^{-1} \cdot 1.83 \\ &= 609.49\end{aligned}$$

  This is a relative reduction of *21%*. Knowledge of conversion has a larger effect.

Multivariate Gaussian

- Propagation of a Gaussian through a linear model yields another Gaussian.
- Consider $X \sim \mathcal{N}(\mu, \Sigma)$ and $Y = \varphi(X) = AX + b$ with $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^n$.
- Can show that $Y \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$.
- Property can be used to establish useful properties.
- For example, consider Gaussian *mixture* $Y = X_1 + X_2$ with independent $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.
- One can show that Gaussian mixture has form $Y = AX$ and that $Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Multivariate Gaussian

- Understanding geometry of multivariate Gaussian facilitates *data visualization*.

- We write the covariance between X_1 and X_2 as:

$$\Sigma_{12} = \rho \cdot \sqrt{\Sigma_{11}\Sigma_{22}}$$

where $\rho = \mathrm{PCorr}[X]_{12}$ is Pearson correlation coefficient.

- With this, joint pdf can be expressed as:

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\Sigma_{11}\Sigma_{22}}} \exp\left(-E(x_1, x_2)\right)$$

where:

$$E(x_1, x_2) = \frac{1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\Sigma_{11}} + \frac{(x_2-\mu_2)^2}{\Sigma_{22}} - 2\rho \frac{(x_1-\mu_1)(x_2-\mu_2)}{\sqrt{\Sigma_{11}\Sigma_{22}}} \right]$$

If we fix $f(x_1, x_2)$ to a given value p, we can write:

$$\ell = E(x_1, x_2)$$

with $\ell = -\log(p \cdot 2\pi\sqrt{\Sigma_{11}\Sigma_{22}})$.

Multivariate Gaussian

$$E(x_1, x_2) = \frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\Sigma_{11}} + \frac{(x_2-\mu_2)^2}{\Sigma_{22}} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sqrt{\Sigma_{11}\Sigma_{22}}}\right]$$

- Equation $\ell = E(x_1, x_2)$ defines an ellipse whose size scales with ℓ.
- Ellipse $\ell = E(x_1, x_2)$ is centered around (μ_1, μ_2).
- Length of axis in x_1 direction is defined by Σ_{11}.
- Length of axis in x_2 direction is defined by Σ_{22}.
- Correlation between X_1 and X_2 defines *orientation of ellipse*:
  - If $\rho > 0$ (positive correlation), ellipse is tilted to right
  - If $\rho < 0$ (negative correlation), ellipse is tilted to left
  - If $\rho = 0$, (no correlation), ellipse has no tilt.
- Correlation *does not* affect size of ellipsoid (only its orientation).

Multivariate Gaussian

- Want *confidence regions* (regions over which we expect to find X with a given prob).

- We want to determine box region $\mathcal{B} \subseteq \mathcal{D}$ that contains X with probability $1-\alpha$:

$$\mathbb{P}(X \in \mathcal{B}) = 1 - \alpha.$$

- For $X \sim \mathcal{N}(\mu, \sigma^2)$ (1D), box is a line interval:

$$\mathcal{B} = \{\, x \in [\mu \pm \sqrt{\mathbb{Q}(1-\alpha)}\sigma]\}$$

  where $\mathbb{Q}(1-\alpha)$ is the $(1-\alpha)$-quantile of $\chi^2(1)$.

- Box tells us that a realization of $\mathcal{N}(\mu, \sigma^2)$ will land in $\mathcal{B}$ with prob $1-\alpha$.

Multivariate Gaussian

- For $X = (X_1, X_2)$ and $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ we can define box:

$$\mathcal{B} = \{\, x_1 \in [\mu_1 \pm \sqrt{\mathbb{Q}(1-\alpha)}\sigma_1] \,\&\, x_2 \in [\mu_2 \pm \sqrt{\mathbb{Q}(1-\alpha)}\sigma_2]\}.$$

- This is known as *marginal box* and does not capture correlations in X_1 and X_2.

Multivariate Gaussian

- For $X \sim \mathcal{N}(\mu, \Sigma)$, we have seen that joint pdf has level sets that form ellipses.
- This means that realizations of X concentrate in ellipses.
- We thus seek an *ellipsoidal* region $\mathcal{E} \subseteq \mathcal{D}$ that contains X with some probability:

$$\mathbb{P}(X \in \mathcal{E}) = 1 - \alpha.$$

- One can show that the ellipsoidal region is given by:

$$\mathcal{E} = \{(x-\mu)^T \Sigma^{-1} (x-\mu) \leq \mathbb{Q}(1-\alpha)\}$$

where $\mathbb{Q}(1-\alpha)$ is the $(1-\alpha)$-quantile of $\chi^2(n)$.

Multivariate Gaussian

- Can show that *tightest box* that encloses ellipsoid $\mathcal{E}$ is:

$$\mathcal{B} = \{\, x_1 \in [\mu_1 \pm \sqrt{\mathbb{Q}(1-\alpha)}\sigma_1] \,\&\, x_2 \in [\mu_2 \pm \sqrt{\mathbb{Q}(1-\alpha)}\sigma_2] \,\}.$$

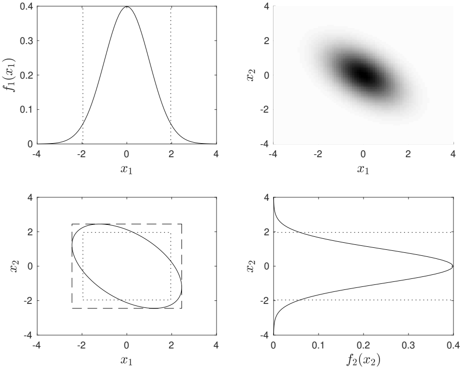
  where $\mathbb{Q}(1-\alpha)$ is $(1-\alpha)$-quantile of $\chi^2(n)$.

- Note difference of quantile with that of marginal box.

- Ellipsoid and bounding box capture correlations between RVs.

Example: Geometry of 2D Gaussian Ellipsoid `ch3_gaussian_geometry.m`

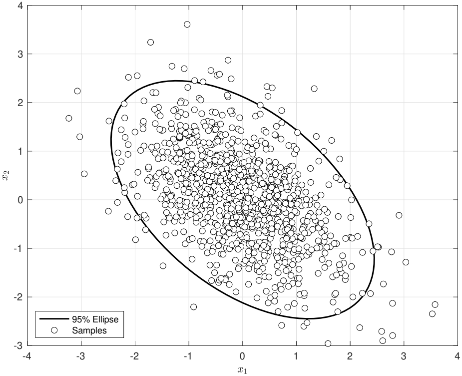
We visualize joint pdf and marginal pdfs for $\mu = (0,0)$ and $\Sigma = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$.



- Ellipsoid is tilted to left (because $\Sigma_{12} < 0$) and is centered at $\mu = (0,0)$.
- Axes along x_1 and x_2 directions have same length (because $\Sigma_{11} = \Sigma_{22}$).
- Marginal box does not capture its shape (its tilt) while the enclosing box does.

Example: Geometry of 2D Gaussian Ellipsoid `ch3_gaussian_samples_ellipse.m`

We now present 95% ellipsoidal region and 1000 realizations from Gaussian.



- We have that 942 out of $\mathcal{S}$ =1000 realizations (94.2%) lie inside ellipsoid
- Realizations inside ellipse satisfy constraint $(x_\omega - \mu)^T \Sigma^{-1} (x_\omega - \mu) \leq \mathbb{Q}(1-\alpha)$
- Fraction of realizations that are inside ellipsoid is not exactly 95%. Why?