Shuyang Ling

New York University Shanghai

September 9, 2019
If $X \sim \mathcal{N}(0, \Sigma)$ where

$$
\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} = U\Lambda U^\top, \quad U = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}
$$

Then

$$U^\top X \sim \mathcal{N}(0, \tilde{\Sigma}), \quad \tilde{\Sigma} = \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix} = \Lambda, \quad U^\top = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$
Marginal distribution of multivariate Gaussian

Let \( e_i = [0, \cdots, 0, 1, 0, \cdots, 0]^\top \in \mathbb{R}^d \), consider \( e_i^\top X \) where \( X \sim \mathcal{N}(\mu, \Sigma) \).

How about its center (mean) and spread (variance)?

\[
\mu_i = e_i^\top \mu, \quad \sigma_i^2 = e_i^\top \Sigma e_i = \sigma_{ii}
\]

where \( \mu_i \) is the \( i \)-th entry of \( \mu \) and \( \sigma_{ii} \) is the \( i \)-th diagonal entry of \( \Sigma \).

The marginal distribution of \( X_1 \) is \( X_1 \sim \mathcal{N}(\mu_i, \sigma_{ii}) \), where \( \sigma_{ii} \) is the \((i, i)\) entry of \( \Sigma \).

The marginal distribution:

\[
X \sim \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}, \quad X_1 \sim \mathcal{N}(0, 1), \quad X_2 \sim \mathcal{N}(0, 1).
\]
Marginal distribution of Gaussian
When \(A\) is a vector

Let \(a \in \mathbb{R}^d\) be a vector and \(Z \sim \mathcal{N}(\mu, \Sigma) \in \mathbb{R}^d\), then

\[
a^\top Z = \sum_{i=1}^{d} a_i Z_i \sim \mathcal{N}(a^\top \mu, a^\top \Sigma a).
\]

(A special case: \(a_i = 1, \forall 1 \leq i \leq d\)).

If \(Z_i \sim \mathcal{N}(\mu_i, \sigma_i^2), 1 \leq i \leq d\) are independent (\(\Sigma\) is diagonal and \(\sigma_{ii} = \sigma_i^2\)), then

\[
\sum_{i=1}^{d} Z_i \sim \mathcal{N}\left(\sum_{i=1}^{d} \mu_i, \sum_{i=1}^{d} \sigma_i^2\right)
\]

where \(a_i\) is the \(i\)-th entry of \(a\).
Return to the previous example:

\[ \mathbf{X} \sim \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \]

What is distribution of \( X_1 + X_2 \)?

\[ 1^T \Sigma 1 = \sigma_{11} + \sigma_{22} + 2\sigma_{12} = 3. \]

Therefore, \( X_1 + X_2 \sim \mathcal{N}(0, 3) \).
Given a random vector \((X, Y)\). The marginal distributions are both Gaussian. Is the joint distribution Gaussian?

**No**, one example is in our homework.

Suppose \(a\) and \(b\) are two deterministic vectors in \(\mathbb{R}^d\).

**Question:** When are \(a^\top Z\) and \(b^\top Z\) independent where \(Z \sim \mathcal{N}(0, I_d)\)?

**Independence**

The independence holds if and only if

\[
a^\top b = \sum_{i=1}^{d} a_i b_i = 0,
\]

i.e., \(a\) and \(b\) are orthogonal to each other.
Given a random variable, we are interested in the average value of its outcome.

**Expectation and mean**

The **expectation** or mean of $X$ is defined to be

- **Discrete**: let $f_X(x)$ be the pmf of $X$,

$$E(X) = \sum_{x \in \mathcal{X}} xf_X(x) = \sum_{x \in \mathcal{X}} x \cdot P(X = x).$$

- **Continuous**: let $f_X(x)$ be the pdf of $X$,

$$E(X) = \int_{\mathbb{R}} xf_X(x) \, dx$$
The mean is a deterministic quantity which describes the center of mass of the distribution.

As a unifying form, we denote

$$E(X) = \int_{\mathbb{R}} x \, dF_X(x), \quad F_X \text{ is the cdf of } X.$$ 

In particular, if $X$ is continuous, $dF_X(x) = f_X(x) \, dx$. 
Example

Suppose $X \sim \text{Bernoulli}(p)$. What is $\mathbb{E}(X)$?

$$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x \mathbb{P}(X = x)$$

$$= 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = p.$$

Let $X \sim \text{Unif}(a, b)$:

$$\mathbb{E}(X) = \int_{a}^{b} \frac{x}{b-a} \, dx = \frac{1}{b-a} \cdot \frac{x^2}{2} \bigg|_{a}^{b}$$

$$= \frac{1}{2(b-a)} (b^2 - a^2) = \frac{a + b}{2}.$$
How about Gaussian distribution? Let $X \sim \mathcal{N}(\mu, \sigma^2)$

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} xe^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx.$$ 

By letting $t = x - \mu$ and $x = t + \mu$, there holds

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} te^{-\frac{t^2}{2\sigma^2}} \, dt + \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \mu e^{-\frac{t^2}{2\sigma^2}} \, dt = \mu.$$ 

The first term vanishes since $te^{-\frac{t^2}{2\sigma^2}}$ is an odd function!
## Mean of important random variables

<table>
<thead>
<tr>
<th>Random variable</th>
<th>Parameters</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>$p$</td>
<td>$p$</td>
</tr>
<tr>
<td>Geometric</td>
<td>$p$</td>
<td>$\frac{1}{p}$</td>
</tr>
<tr>
<td>Binomial</td>
<td>$n, p$</td>
<td>$np$</td>
</tr>
<tr>
<td>Poisson</td>
<td>$\lambda$</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>Uniform</td>
<td>$a, b$</td>
<td>$\frac{a+b}{2}$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\beta$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\mu, \sigma$</td>
<td>$\mu$</td>
</tr>
</tbody>
</table>
Existence of expectation

Not every distribution has an expected value.

We say the expectation exists if

$$\mathbb{E}(|X|) = \int_{\mathbb{R}} |x| \, dF_X(x) < \infty.$$ 

In particular, 

- If $X$ is discrete,

  $$\sum_{x \in \mathcal{X}} |x| f_X(x) < \infty$$

- If $X$ is continuous,

  $$\int_{\mathcal{X}} |x| f_X(x) \, dx < \infty$$
Example

One famous example is Cauchy distribution with pdf

\[ f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}. \]

We proceed to compute \( \mathbb{E}(|X|) \):

\[
\int_{\mathbb{R}} |x| \, dF(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1 + x^2} \, dx \\
= \frac{2}{\pi} \int_{0}^{\infty} x \, dx \\
= \frac{1}{\pi} \int_{0}^{\infty} \frac{d(x^2)}{1 + x^2} \\
= \frac{1}{\pi} \ln(1 + x^2) \bigg|_0^\infty = \infty.
\]
A gambling game: toss a fair die and win/lose money depending on the outcome

<table>
<thead>
<tr>
<th>Outcome X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profit $r(X)$</td>
<td>-5</td>
<td>-2</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The average profits after one toss would be

$$\text{Average profits} = \sum_{x=1}^{6} \mathbb{P}(X = x) \cdot r(x) = \frac{1}{6} \sum_{x=1}^{6} r(x)$$

$$= \frac{1}{6} \cdot (-5 - 2 - 1 + 1 + 3 + 4)$$

$$= 0.$$ 

The expected return should be the sum of the return $r(x)$ times the chance that $X = x$ over all values of $x$. 
Expectation of $r(X)$

- **Discrete scenario:** $f_X(x)$ is a pmf and
  \[
  \mathbb{E}(r(X)) = \sum_{x \in \mathcal{X}} r(x)f_X(x).
  \]

- **Continuous scenario:** $f_X(x)$ is a pdf and
  \[
  \mathbb{E}(r(X)) = \int_{x \in \mathcal{X}} r(x)f_X(x) \, dx.
  \]
Expectation of $r(X, Y)$

This can be easily extended to multivariate case.

- Suppose we have two random variables $(X, Y)$ with joint pmf/pdf $f_{X,Y}(x, y)$.
- What’s the expectation of $r(X, Y)$ where $r$ is a function on $\mathbb{R}^2$?

Expectation of $r(X)$

- Discrete scenario: $f_X(x, y)$ is a pmf and

$$\mathbb{E}(r(X, Y)) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} r(x, y) f_{X,Y}(x, y).$$

- Continuous scenario: $f_X(x)$ is a pdf and

$$\mathbb{E}(r(X, Y)) = \int_{x \in \mathcal{X}, y \in \mathcal{Y}} r(X, Y) f_{X,Y}(x, y) \, dx \, dy.$$

This can be generalized to $n$ variables.
Suppose we have two random variables. What is the expectation of $X + Y$?

**Linearity**

Let $X$ and $Y$ be two random variables (not necessarily independent).

\[
\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)
\]

\[
\mathbb{E}(cX) = c \mathbb{E}(X)
\]

where $c$ is a scalar.
Linearity of expectation

We show that for discrete random variables,

\[ E(X + Y) = \sum_{x \in X, y \in Y} (x + y) f_{X,Y}(x, y) \]

\[ = \sum_{x \in X} \sum_{y \in Y} xf_{X,Y}(x, y) + \sum_{y \in Y} \sum_{x \in X} yf_{X,Y}(x, y) \]

\[ = \sum_{x \in X} xf_{X}(x) + \sum_{y \in Y} yf_{Y}(y) \]

\[ = E(X) + E(Y). \]

The derivation does NOT require independence.
Suppose $X \sim \text{Binomial}(n, p)$. What is $\mathbb{E}(X)$?

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \cdot \mathbb{P}(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k}.$$ 

Note that $X = \sum_{i=1}^{n} X_i$ where each $X_i \sim \text{Bernoulli}(p)$. Thus

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} p = np.$$
Let $A$ be a set and define

$$1_A(X) = \begin{cases} 1, & \text{if } X \in A \\ 0, & \text{otherwise} \end{cases}$$

This function is called the indicator function of $A$.

$$\mathbb{E} 1_A(X) = \int_{x \in \mathcal{X}} 1_A(x) \, dF_X(x) = \int_A dF_X(x) = \mathbb{P}(X \in A).$$

In other words, the probability of an event occurring is a special case of expectation.
X and Y are independent if and only if for any set $A$ and $B$,

$$
\mathbb{E} 1_A(X)1_B(Y) = \mathbb{E} 1_A(X) \mathbb{E} 1_B(Y)
$$

since $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$.

- If $X$ and $Y$ are independent, then

$$
\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y).
$$

- If $X$ and $Y$ are independent, then

$$
\mathbb{E} g_1(X)g_2(Y) = \mathbb{E} g_1(X) \mathbb{E} g_2(Y)
$$

for any piecewise continuous functions $g_1$ and $g_2$. 
An important example of $r(x)$ is $r(x) = (x - a)^2$ for some number $a$.

**Variance**

Let $X$ be a random variable with mean $\mu$. The variance, denoted by $\sigma^2$ or $\sigma_X^2$ or $\text{Var}(X)$, is defined by

$$\sigma^2 = \mathbb{E}(X - \mu)^2 = \int_X (x - \mu)^2 \, dF(x)$$

provided this expectation exists. The standard deviation is $\sigma$.

- Variance, the expectation of the squared deviation of a random variable from its mean, is a way to mean the spread of a distribution (uncertainty) from its mean.
- $\sigma^2$ is always nonnegative. In particular, if $\sigma = 0$, then $X = \mu$. 