4 Jointly distributed random variables

1. Introduction
2. Independent Random Variables
3. Transformations
4. Covariance, Correlation
5. Conditional Distributions
6. Bivariate Normal Distribution
4.1 Introduction

Computation of probabilities with more than one random variable

- two random variables ... bivariate
- two or more random variables ... multivariate

Concepts:

- Joint distribution function
- purely discrete: Joint probability function
- purely continuous: Joint density
Joint distribution function

Bivariate case: Random variables $X$ and $Y$

Define joint distribution function as

$$F(x, y) := P(X \leq x, Y \leq y), \quad -\infty < x, y < \infty$$

Bivariate distribution thus fully specified

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

for $x_1 < x_2$ and $y_1 < y_2$

Marginal distribution: $F_X(x) := P(X \leq x) = F(x, \infty)$

Idea: $P(X \leq x) = P(X \leq x, Y < \infty) = \lim_{y \to \infty} F(x, y)$

Analogous $F_Y(y) := P(Y \leq y) = F(\infty, y)$
Bivariate continuous random variable

$X$ and $Y$ jointly continuous if there exists joint density function:

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

Joint distribution function then obtained by integration

$$F(a, b) = \int_{y=-\infty}^{b} \int_{x=-\infty}^{a} f(x, y) \, dx \, dy$$

Density of marginal distribution $X$ obtained by integration over $\Omega_Y$:

$$f_X(x) = \int_{y=-\infty}^{\infty} f(x, y) \, dy$$
Example: Bivariate uniform distribution

$X$ and $Y$ uniformly distributed on $[0,1] \times [0,1]$ ⇒ density $$f(x, y) = 1, \quad 0 \leq x, y \leq 1.$$  

Joint distribution function

$$F(a, b) = \int_{y=0}^{b} \int_{x=0}^{a} f(x, y) \, dx \, dy = a \, b, \quad 0 \leq a, b \leq 1.$$  

Density of marginal distribution:

$$f_X(x) = \int_{y=-\infty}^{\infty} f(x, y) \, dy = 1, \quad 0 \leq x \leq 1$$  

i.e. density of univariate uniform distribution
Exercise: Bivariate continuous distribution

Let joint density of $X$ and $Y$ be given as

$$f(x, y) = \begin{cases} 
ax^2y^2, & 0 < x < y < 1 \\
0, & \text{else}
\end{cases}$$

1. Determine the constant $a$

2. Determine the CDF $F(x, y)$

3. Compute the probability $P(X < 1/2, Y > 2/3)$

4. Compute the probability $P(X < 2, Y > 2/3)$

5. Compute the probability $P(X < 2, Y > 3)$

**Hint:** Be careful with the domain of integration
Let joint density of $X$ and $Y$ be given as

$$f(x, y) = \begin{cases} 
  e^{-(x+y)}, & 0 < x < \infty, 0 < y < \infty \\
  0, & \text{else}
\end{cases}$$

Find the density function of the random variable $X/Y$

Again most important to consider the correct domain of integration

$$P(X/Y \leq a) = \int_{y=0}^{\infty} \int_{x=0}^{ay} e^{-(x+y)} \, dx \, dy = 1 - \frac{1}{a + 1}$$

(More details are given in the book)
Bivariate discrete random variable

$X$ and $Y$ both discrete

Define joint probability function

$$p(x, y) = P(X = x, Y = y)$$

Naturally we have

$$p(x, y) = F(x, y) - F(x^-, y) - F(x, y^-) + F(x^-, y^-)$$

Obtain probability function of $X$ by summation over $\Omega_Y$:

$$p_X(x) = P(X = x) = \sum_{\Omega_Y} p(x, y)$$
Example

Bowl with 3 red, 4 white and 5 blue balls; draw 3 balls randomly without replacement

\(X \ldots \) number of red drawn balls

\(Y \ldots \) number of white drawn balls

e. g.: \(p(0, 1) = P(0R, 1W, 2B) = \frac{\binom{3}{0} \binom{4}{1} \binom{5}{2}}{\binom{12}{3}} = \frac{40}{220}\)

<table>
<thead>
<tr>
<th>(i)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>(p_X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10/220</td>
<td>40/220</td>
<td>30/220</td>
<td>4/220</td>
<td>84/220</td>
</tr>
<tr>
<td>1</td>
<td>30/220</td>
<td>60/220</td>
<td>18/220</td>
<td>0</td>
<td>108/220</td>
</tr>
<tr>
<td>2</td>
<td>15/220</td>
<td>12/220</td>
<td>0</td>
<td>0</td>
<td>27/220</td>
</tr>
<tr>
<td>3</td>
<td>1/220</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/220</td>
</tr>
</tbody>
</table>

| \(p_Y\) | 56/220 | 112/220 | 48/220 | 4/220 | 1       |
**Multivariate random variable**

More than two random variables

joint distribution function for \( n \) random variables

\[
F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)
\]

Discrete: Joint probability function:

\[
p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)
\]

Marginals again by summation over all components that are not of interest, e.g.

\[
p_{X_1}(x_1) = \sum_{x_2 \in \Omega_2} \cdots \sum_{x_n \in \Omega_n} p(x_1, \ldots, x_n)
\]
Multinomial distribution

one of the most important multivariate discrete distribution

$n$ independent experiments with $r$ possible results, each with probability $p_1, \ldots, p_r$

$X_i$ ... number of experiments with $i$th result, then

$$P(X_1 = n_1, \ldots, X_r = n_r) = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r}$$

whenever $\sum_{i=1}^{r} n_i = n$

Generalization of binomial distribution ($r = 2$)

Exercise: Poker Dice (throw 5 dices)

Compute probability of a (High) straight, Four of a kind, Full House
4.2 Independent random variables

Two random variables $X$ and $Y$ are called independent if for all events $A$ and $B$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Information on $X$ does not give any information on $Y$

$X$ and $Y$ independent if and only if

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$$

i.e. $F(a, b) = F_X(a) F_Y(b)$ for all $a, b$.

Also equivalent with $f(x, y) = f_X(x) f_Y(y)$ in the continuous case and with $p(x, y) = p_X(x) p_Y(y)$ in the discrete case for all $x, y$. 
Example: continuous

Looking back at the example

\[ f(x, y) = \begin{cases} 
ax^2 y^2, & 0 < x < y < 1 \\
0, & \text{else}
\end{cases} \]

Are the random variables \(X\) and \(Y\) with joint density as specified above independent?

What changes if we look at

\[ f(x, y) = \begin{cases} 
ax^2 y^2, & 0 < x < 1, 0 < y < 1 \\
0, & \text{else}
\end{cases} \]
Example

$Z \sim \mathcal{P} (\lambda)$ \ldots number of persons which enter a bar

$p$ \ldots percentage of male visitors

$X, Y$ \ldots number of male and female visitors respectively

$X, Y$ are independently Poisson distributed with parameters $p\lambda$ and $q\lambda$ respectively

Solution: Law of total probability:

$$P(X = i, Y = j) = P(X = i, Y = j | X + Y = i + j) P(X + Y = i + j)$$

by definition:

$$P(X = i, Y = j | X + Y = i + j) = \binom{i+j}{i} p^i q^j$$

$$P(X + Y = i + j) = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

Together:

$$P(X = i, Y = j) = e^{-\lambda} \frac{(\lambda p)^i}{i!} \frac{(\lambda q)^j}{j!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} e^{-\lambda q} \frac{(\lambda q)^j}{j!}$$
Example: Two dices

$X, Y \ldots$ uniformly distributed on $\{1, \ldots, 6\}$

Due to independence we have $p(x, y) = p_X(x) p_Y(y) = \frac{1}{36}$

Distribution function:

$F_X(x) = F_Y(x) = \frac{i}{6}, \text{ if } x \leq i < x + 1 \text{ and } 0 < x < 7$

$F(x, y) = F_X(x)F_Y(y) = \frac{ij}{36}, \text{ if } x \leq i < x + 1, y \leq j < y + 1$

Which distribution has $X + Y$?

$P(X + Y = 2) = p(1, 1) = 1/36$

$P(X + Y = 3) = p(1, 2) + p(2, 1) = 2/36$

$P(X + Y = 4) = p(1, 3) + p(2, 2) + p(3, 1) = 3/36$

$P(X + Y = k) = p(1, k - 1) + p(2, k - 2) + \cdots + p(k - 1, 1)$
**Sum of two independent distributions**

Sum of random variables itself a random variable

Computation of distribution via convolution

Discrete random variables:

\[
P(X + Y = k) = \sum_{x+y=k} p_X(x)p_Y(y) = \sum_{\Omega_Y} p_X(k-y)p_Y(y)
\]

Continuous random variables:

\[
f_{X+Y}(x) = \int_{y=-\infty}^{\infty} f_X(x-y)f_Y(y)dy
\]

**Exercise:** \(X \sim \mathcal{P}(\lambda_1)\) and \(Y \sim \mathcal{P}(\lambda_2)\) independent

\[\Rightarrow \quad X + Y \sim \mathcal{P}(\lambda_1 + \lambda_1)\]
Example of convolution: Continuous case

$X$, $Y$ independent, uniformly distributed on $[0, 1]$

i.e. $f(x, y) = 1$, $(x, y) \in [0, 1] \times [0, 1]$

$f_X(x) = 1$, $0 \leq x \leq 1$, $f_Y(y) = 1$, $0 \leq y \leq 1$

Computation of density of $Z := X + Y$

$$f_Z(x) = \int_{y = -\infty}^{\infty} f_X(x - y) f_Y(y) dy$$

$$= \begin{cases} 
\int_{y=0}^{x} dy = x, & 0 < x \leq 1 \\
\int_{y=x-1}^{1} dy = 2 - x, & 1 < x \leq 2 
\end{cases}$$

Reason: $f_Y(y) = 1$ if $0 \leq y \leq 1$

$f_X(x - y) = 1$ if $0 \leq x - y \leq 1$ $\iff$ $y \leq x \leq y + 1$
Another example of convolution

$X$, $Y$ indep., $\Gamma-$distributed with parameters $t_1, t_2$ and same $\lambda$

$$f_X(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{t_1-1}}{\Gamma(t_1)}, \quad f_Y(y) = \frac{\lambda e^{-\lambda y}(\lambda y)^{t_2-1}}{\Gamma(t_2)}, \quad x, y \geq 0,$$

$$f_Z(x) = \int_{y=-\infty}^{\infty} f_X(x-y) f_Y(y) dy$$

$$= \int_{y=0}^{x} \frac{\lambda e^{-\lambda(x-y)}(\lambda(x-y))^{t_1-1}}{\Gamma(t_1)} \frac{\lambda e^{-\lambda y}(\lambda y)^{t_2-1}}{\Gamma(t_2)} dy$$

$$= \frac{\lambda^{t_1+t_2} e^{-\lambda x}}{\Gamma(t_1) \Gamma(t_2)} \int_{y=0}^{x} (x-y)^{t_1-1} y^{t_2-1} dy$$

$$= \left| \begin{array}{ll} y & = xz \\ dy & = xdz \end{array} \right| = \frac{\lambda e^{-\lambda x}(\lambda x)^{t_1+t_2-1}}{\Gamma(t_1 + t_2)}$$
4.3 Transformations

As in the univariate case one considers also in the multivariate case transformations of random variables.

We restrict ourselves to the bivariate case.

$X_1$ and $X_2$ jointly distributed with density $f_{X_1,X_2}$

$Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$

**Major question:** What is the joint density $f_{Y_1,Y_2}$?
Density of Transformation

Assumptions:

- \( y_1 = g_1(x_1, x_2) \) and \( y_2 = g_2(x_1, x_2) \) can be uniquely solved for \( x_1 \) and \( x_2 \), say \( x_1 = h_1(y_1, y_2) \) and \( x_2 = h_2(y_1, y_2) \)

- \( g_1 \) and \( g_2 \) are \( C^1 \) such that

\[
J(x_1, x_2) = \begin{vmatrix}
\frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\
\frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2}
\end{vmatrix} \neq 0
\]

Under these conditions we have

\[
f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2) |J(x_1, x_2)|^{-1}
\]

Proof: Calculus

Note the similarity to the one-dimensional case
Examples

• Sheldon Ross: Example 7a
  \[ Y_1 = X_1 + X_2 \]
  \[ Y_2 = X_1 - X_2 \]

• Sheldon Ross: Example 7c
  \( X \) and \( Y \) independent gamma random variables
  \[ U = X + Y \]
  \[ V = X/(X + Y) \]
Expectation of bivariate RV

$X$ and $Y$ discrete with joint probability $p(x, y)$.

Like in the one-dimensional case:

$$E(g(X, Y)) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} g(x, y)p(x, y)$$

Exercise:

Let $X$ and $Y$ eyes of two fair dice, independently thrown

Compute expectation of the difference $|X - Y|$
Expectation of bivariate RV

\(X\) and \(Y\) continuous with joint density \(f(x, y)\).

Like in the one-dimensional case:

\[
E(g(X, Y)) = \int_{x \in \mathbb{R}} \int_{y \in \mathbb{R}} g(x, y) f(x, y) \, dy \, dx
\]

Exercise:

Accident occurs at point \(X\) on a road of length \(L\). At that time ambulance is at point \(Y\). Assuming both \(X\) and \(Y\) uniformly distributed on \([0, L]\) and independent, calculate average distance between them. This is again \(E(|X - Y|)\)
Expectation of the sum of two RV

$X$ and $Y$ discrete with joint probability $p(x, y)$

For $g(x, y) = x + y$ we obtain

$$E(X + Y) = \sum_{x \in \Omega_X} \sum_{y \in \Omega_Y} (x + y)p(x, y) = E(X) + E(Y)$$

Analogous in the continuous case:

$$E(X + Y) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} (x + y)f(x, y) \, dx \, dy = E(X) + E(Y)$$

**Be aware:** Additivity for variances in general not the case!
Expectation of functions under independence

If \( X \) and \( Y \) are independent random variables, then for any functions \( h \) and \( g \) we obtain

\[
E \left( g(X)h(Y) \right) = E \left( g(X) \right) E \left( h(Y) \right)
\]

Proof:

\[
E \left( g(X)h(Y) \right) = \int \int g(x)h(y)f(x, y) \, dx \, dy
\]

\[
= \int \int g(x)h(y)f_X(x)f_Y(y) \, dx \, dy
\]

\[
= \int g(x)f_X(x) \, dx \int h(y)f_Y(y) \, dy
\]

\[
= E \left( g(X) \right) E \left( h(Y) \right)
\]

Second equality uses independence
Expectation of random samples

\( X_1, \ldots, X_n \) i.i.d. like \( X \), (independent, identically distributed)

**Definition:** \( \bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i \)

For \( E(X) = \mu \) and \( \text{Var}(X) = \sigma^2 \) we obtain:

\[
E(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}
\]

**Proof:**

\[
E \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} E(X_i)
\]

\[
E \left( \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right)^2 = \frac{1}{n^2} E \left( \sum_{i=1}^{n} (X_i - \mu) \right)^2 = \frac{1}{n^2} E \left( \sum_{i=1}^{n} (X_i - \mu)^2 \right)
\]

Last equality due to independence of \( X_i \)
4.4 Covariance and Correlation

Describe the relation between two random variables

Definition Covariance:

\[ \text{Cov}(X, Y) = E(X - E(X))(Y - E(Y)) \]

Usual notation: \( \sigma_{XY} := \text{Cov}(X, Y) \)

Just like for variances we have

\[ \sigma_{XY} = E(XY) - E(X)E(Y) \]

Definition of correlation:

\[ \rho(X, Y) := \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \]
Example Correlation

\[ \rho = 0.9 \]

\[ \rho = -0.6 \]

\[ \rho = 0.3 \]

\[ \rho = 0.0 \]
Example Covariance

Discrete bivariate distribution ($\Omega_X = \Omega_Y = \{0, 1, 2, 3\}$) given by

<table>
<thead>
<tr>
<th>$i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>$p_X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/20</td>
<td>4/20</td>
<td>3/20</td>
<td>2/20</td>
<td>10/20</td>
</tr>
<tr>
<td>1</td>
<td>3/20</td>
<td>2/20</td>
<td>2/20</td>
<td>0</td>
<td>7/20</td>
</tr>
<tr>
<td>2</td>
<td>1/20</td>
<td>1/20</td>
<td>0</td>
<td>0</td>
<td>2/20</td>
</tr>
<tr>
<td>3</td>
<td>1/20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/20</td>
</tr>
</tbody>
</table>

$p_Y$ 6/20 7/20 5/20 2/20 20/20

Compute Cov ($X$, $Y$)

Solution:

$\text{Cov} (X, Y) = E(XY) - E(X)E(Y) = \frac{8}{20} - \frac{14}{20} \cdot \frac{23}{20} = -\frac{162}{400}$
Example 2: Covariance:

Compute covariance between $X$ and $Y$ for

$$f(x, y) = \begin{cases} 
24xy, & 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\
0, & \text{else}
\end{cases}$$

Compute first cdf of $X$:

$$P(X \leq a) = \int_{y=0}^{1-a} \int_{x=0}^{a} 24xy \, dx \, dy + \int_{y=1-a}^{1} \int_{x=0}^{1-y} 24xy \, dx \, dy$$

$$= 6(1-a)^2a^2 + 1 - 6(1-a)^2 + 8(1-a)^3 - 3(1-a)^4$$

and by differentiation

$$f_X(x) = 12(1-x)^2x - 12(1-x)x^2 + 12(1-x) - 24(1-x)^2 + 12(1-x)^3$$

$$= 12(1-x)^2x$$
Example 2: Covariance continued

Because of symmetry we have \( f_Y(x) = f_X(x) \) and we get

\[
E(X) = E(Y) = \int_{x=0}^{1} 12(1-x)^2x^2 \, dx = 2/5
\]

Furthermore

\[
E(XY) = \int_{y=0}^{1} \int_{x=0}^{1-y} 24x^2y^2 \, dx \, dy = 2/15
\]

and finally

\[
\text{Cov} (X, Y) = \frac{2}{15} - \frac{2}{5} \cdot \frac{2}{5} = \frac{10 - 12}{75} = -\frac{2}{75}
\]
Covariance for independent RV

\( X \) and \( Y \) independent \( \Rightarrow \sigma_{XY} = 0 \)

follows immediately from \( \sigma_{XY} = E(XY) - E(X)E(Y) \) and

\[
E(XY) = \sum_x \sum_y xy p(x,y) = \sum_x x p_X(x) \sum_y y p_Y(y)
\]

Converse not true:

\( X \) uniformly distributed on \( \{-1, 0, 1\} \) and \( Y = \begin{cases} 
0, & X \neq 0 \\
1, & X = 0 
\end{cases} \)

\[
E(X) = 0 \\
XY = 0 \Rightarrow E(XY) = 0
\]

thus \( \text{Cov} (X, Y) = 0 \), although \( X \) and \( Y \) not independent:

e.g. \( P(X = 1, Y = 0) = P(X = 1) = 1/3, P(Y = 0) = 2/3 \)
Properties of Covariance

Obviously

\[ \text{Cov}(X, Y) = \text{Cov}(Y, X), \quad \text{and} \quad \text{Cov}(X, X) = \text{Var}(X) \]

Covariance is a bilinear form:

\[ \text{Cov}(aX, Y) = a \text{Cov}(X, Y), \quad a \in \mathbb{R} \]

and

\[
\text{Cov} \left( \sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \text{Cov}(X_i, Y_j)
\]

Proof by simple computation …
**Variance of a sum of RV**

Due to the properties shown above

\[
\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov} \left( X_i, X_j \right)
\]

\[
= \sum_{i=1}^{n} \text{Var} \left( X_i \right) + \sum_{i=1}^{n} \sum_{j \neq i} \text{Cov} \left( X_i, X_j \right)
\]

**Extreme cases:**

- **independent RV:** \( \text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var} \left( X_i \right) \)

- \( X_1 = X_2 = \cdots = X_n \): \( \text{Var} \left( \sum_{i=1}^{n} X_i \right) = n^2 \text{Var} \left( X_1 \right) \)
Correlation

Definition: \( \rho(X, Y) := \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \)

We have:

\[-1 \leq \rho(X, Y) \leq 1\]

Proof:

\[
0 \leq \text{Var} \left( \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} \right) = \frac{\text{Var} (X)}{\sigma_X^2} + \frac{\text{Var} (Y)}{\sigma_Y^2} + \frac{2 \text{Cov} (X, Y)}{\sigma_X \sigma_Y} \\
= 2[1 + \rho(X, Y)]
\]

\[
0 \leq \text{Var} \left( \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \right) = \frac{\text{Var} (X)}{\sigma_X^2} + \frac{\text{Var} (Y)}{\sigma_Y^2} - \frac{2 \text{Cov} (X, Y)}{\sigma_X \sigma_Y} \\
= 2[1 - \rho(X, Y)]
\]
Exercise Correlation

Let $X$ and $Y$ be independently uniform on $[0, 1]$

Compute correlation between $X$ and $Z$ for

1. $Z = X + Y$
2. $Z = X^2 + Y^2$
3. $Z = (X + Y)^2$
4.5 Conditional Distribution

Conditional probability for two events $A$ and $B$:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Corresponding definition for random variables $X$ and $Y$

Discrete:

$$p_{X|Y}(x|y) := P(X = x|Y = y) = \frac{p(x,y)}{p_Y(y)}$$

**Exercise:** Let $p(x,y)$ given by

$$p(0,0) = 0.4, \quad p(0,1) = 0.2, \quad p(1,0) = 0.1, \quad p(1,1) = 0.3,$$

Compute conditional probability function of $X$ for $Y = 1$
Discrete conditional distribution

Conditional CDF:

\[ F_{X|Y}(x|y) := P(X \leq x|Y = y) = \sum_{k \leq x} p_{X|Y}(k|y) \]

If \( X \) and \( Y \) are independent – then \( p_{X|Y}(x|y) = p_X(x) \)

**Proof:** Easy computation

**Exercise:** Let \( X \sim \mathcal{P}(\lambda_1) \) and \( Y \sim \mathcal{P}(\lambda_2) \) be independent.

Compute conditional distribution of \( X \), for \( X + Y = n \)

\[ P(X = k|X + Y = n) = \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)}, \]

\( X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2) \) \( \Rightarrow \) \( X|(X + Y = n) \sim \mathcal{B}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right) \)
Continuous conditional distribution

Continuous: \[ f_{X|Y}(x|y) := \frac{f(x,y)}{f_Y(y)} \quad \text{for } f_Y(y) > 0 \]

Definition in continuous case can be motivated by discrete case (Probabilities for small environments of \(x\) and \(y\))

Computation of conditional probabilities:

\[
P(X \in A|Y = y) = \int_A f_{X|Y}(x|y) \, dx
\]

Conditional CDF:

\[
F_{X|Y}(a|y) := P(X \in (-\infty, a)|Y = y) = \int_{x=-\infty}^{a} f_{X|Y}(x|y) \, dx
\]
Example

Joint density of $X$ and $Y$ given by

$$
 f(x, y) = \begin{cases} 
  c \; x(2 - x - y), & x \in [0, 1], y \in [0, 1], \\
  0, & \text{otherwise.}
\end{cases}
$$

Compute $c$, $f_{X|Y}(x|y)$ and $P(X < 1/2|Y = 1/3)$

Solution: $f_Y(y) = c \int_{x=0}^{1} x(2 - x - y) \, dx = c \left( \frac{2}{3} - \frac{y}{2} \right)$

$$
 1 = \int_{y=0}^{1} f_Y(y) = c \left( \frac{2}{3} - \frac{1}{4} \right) \quad \Rightarrow \quad c = \frac{12}{5}
$$

$$
f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{x(2-x-y)}{\frac{2}{3} - \frac{y}{2}} = \frac{6x(2-x-y)}{4-3y}
$$

$$
P(X < 1/2|Y = 1/3) = \int_{x=0}^{1/2} \frac{6x(2-x-1/3)}{4-3/3} \, dx = \cdots = 1/3
$$
Conditional expectation and variance

Computation in continuous case with conditional density:

\[
E(X|Y = y) = \int_{x=-\infty}^{\infty} x f_{X|Y}(x|y) \, dx
\]

\[
\text{Var}(X|Y = y) = \int_{x=-\infty}^{\infty} (x - E(X|Y = y))^2 f_{X|Y}(x|y) \, dx
\]

**Example:** prolonged

\[
E(X|Y = y) = \int_{x=0}^{1} \frac{6x^2(2 - x - y)}{4 - 3y} \, dx = \frac{5/2 - 2y}{4 - 3y}
\]

Specifically: 
\[
E(X|Y = 1/3) = \frac{2}{9}
\]

\[
\text{Var}(X|Y = y) = \ldots
\]
Conditional expectation and variance

Computation in discrete case with conditional probability function:

\[
E(X|Y = y) = \sum_{x \in \Omega_X} xp_{X|Y}(x|y)
\]

\[
\text{Var} (X|Y = y) = \sum_{x \in \Omega_X} (x - E(X|Y = y))^2 p_{X|Y}(x|y)
\]

**Exercise:** Compute expectation and variance for \(X\) given \(Y = j\)

<table>
<thead>
<tr>
<th>(i)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>(p_X)</th>
<th>(p_Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1/20</td>
<td>4/20</td>
<td>3/20</td>
<td>2/20</td>
<td>10/20</td>
<td>6/20</td>
</tr>
<tr>
<td>1</td>
<td>3/20</td>
<td>2/20</td>
<td>2/20</td>
<td>0</td>
<td>7/20</td>
<td>7/20</td>
</tr>
<tr>
<td>2</td>
<td>1/20</td>
<td>1/20</td>
<td>0</td>
<td>0</td>
<td>2/20</td>
<td>5/20</td>
</tr>
<tr>
<td>3</td>
<td>1/20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/20</td>
<td>2/20</td>
</tr>
<tr>
<td>(p_Y)</td>
<td>6/20</td>
<td>7/20</td>
<td>5/20</td>
<td>2/20</td>
<td>20/20</td>
<td></td>
</tr>
</tbody>
</table>
Computation of expectation by conditioning

$E(X|Y = y)$ can be seen as a function of $y$, and therefore the expectation of this function can be computed.

It holds: $$E(X) = E(E(X|Y))$$

**Proof:** (for the continuous case)

$$E(E(X|Y)) = \int_{y=-\infty}^{\infty} E(X|Y = y) f_Y(y) \, dy$$

$$= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} x f_{X|Y=y}(x) f_Y(y) \, dx \, dy$$

$$= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} x \frac{f(x, y)}{f_Y(y)} f_Y(y) \, dx \, dy = E(X)$$

Verify formula for previous examples.
The conditional variance formula

$$\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E(X|Y))$$

Proof: From $\text{Var}(X|Y) = E(X^2|Y) - (E(X|Y))^2$ we get

$$E(\text{Var}(X|Y)) = E(E(X^2|Y)) - E((E(X|Y))^2) = E(X^2) - E(E(X|Y)^2)$$

On the other hand

$$\text{Var}(E(X|Y)) = E(E(X|Y)^2) - (E(E(X|Y)))^2 = E(E(X|Y)^2) - E(X)^2$$

The sum of both expressions gives the result.

Formula important in the theory of linear regression!
4.6 Bivariate normal distribution

Univariate normal distribution: \[ f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

Standard normal distribution: \[ \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \]

Let \( X_1 \) and \( X_2 \) be independent, both normal distributed \( \mathcal{N}(\mu_i, \sigma_i^2), i = 1, 2 \)

\[
\Rightarrow f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2} - \frac{(x_2-\mu_2)^2}{2\sigma_2^2}}
\]

\[
= \frac{1}{2\pi |\Sigma|^{1/2}} e^{-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}}
\]

where \( x = (x_1, x_2) \), \( \mu = (\mu_1, \mu_2) \), \( \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \)
Density function in general

\( X = (X_1, X_2) \) bivariate normal distributed if joint density has the form

\[
f(x) = \frac{1}{2\pi |\Sigma|^{1/2}} e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)/2}
\]

Covariance matrix: \( \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \)

Notation: \( \rho := \frac{\sigma_{12}}{\sigma_1 \sigma_2} \)

- \( |\Sigma| = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2) \)

- \( \Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix} \)
Density in general

Finally we obtain

$$f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{2(1 - \rho^2)} \right\}$$

where $z_1 = \frac{x_1 - \mu_1}{\sigma_1}$ and $z_2 = \frac{x_2 - \mu_2}{\sigma_2}$ (compare standardization)

Notation suggests that $\mu_i$ and $\sigma_i^2$ are expectation and variance of the marginals $X_i$, and that $\rho$ is the correlation between $X_1$ and $X_2$

Important formula:

$$f(x_1, x_2) = \frac{1}{\sqrt{2\pi \sigma_1}} e^{-\frac{z_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi(1 - \rho^2) \sigma_2}} e^{-\frac{(\rho z_1 - z_2)^2}{2(1 - \rho^2)}}$$

Completion of square in the exponent
Bivariate normal distribution plot

Marginals standard normal distributed $\mathcal{N}(0, 1)$, $\rho = 0$: 
Example density

Let \((X, Y)\) be bivariate normal distributed with \(\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1\) and \(\rho = 1/2\), compute the joint density

**Solution:** \(\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}\)

\(|\Sigma| = 1 - 1/4 = 3/4, \Sigma^{-1} = \frac{4}{3} \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}\)

\((x, y)\Sigma^{-1}(x, y) = \frac{2}{3} (x, y) \begin{pmatrix} 2x-y \\ -x+2y \end{pmatrix} = \frac{4}{3} (x^2 - xy + y^2)\)

\(f(x, y) = \frac{1}{\sqrt{3\pi}} e^{-\frac{2}{3} (x^2 - xy + y^2)}\)

Density after completion of square in exponent:

\(f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \frac{1}{\sqrt{2\pi \cdot 3/4}} e^{-\frac{(y-x/2)^2}{2\cdot3/4}}\)
Example continued

\[ f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \cdot \frac{1}{\sqrt{2\pi} \, 3/4} e^{-\frac{(y-x/2)^2}{2 \cdot 3/4}} \]

Joint density is the product of a standard normal (in \(x\)) and a normal distribution (in \(y\)) with mean \(x/2\) and variance \(3/4\).

Compute density of \(X\):

\[ f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \int_{y=-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \, 3/4} e^{-\frac{(y-x/2)^2}{2 \cdot 3/4}} \, dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \]

\(f_X(x)\) is density of a standard normal distribution

The integral equals 1, because we integrate a density function!
Interpretation of $\mu_i$, $\sigma^2_i$ and $\rho$

In general we have for a bivariate normal distribution

1. $X_1 \sim \mathcal{N}(\mu_1, \sigma^2_1)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma^2_2)$

2. Correlation coefficient $\rho(X_1, X_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$

Proof: 1. Use formula with completed square in exponent and integrate:

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z_1^2}{2}} \int_{x_2=-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} e^{-\frac{(\rho z_1 - z_2)^2}{2(1-\rho^2)}} \, dx_2$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z_1^2}{2}} \int_{s=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\frac{\rho z_1}{\sqrt{1-\rho^2}} - s)^2}{2}} \, ds = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z_1^2}{2}}$$

With substitution $s \leftarrow z_2/\sqrt{1-\rho^2} = (x_2 - \mu_2)/(\sqrt{1-\rho^2}\sigma_2)$
Proof continued

2. Again formula with completed square in exponent and substitutions $z_1 \leftarrow (x_1 - \mu_1)/\sigma_1$, $z_2 \leftarrow (x_2 - \mu_2)/\sigma_2$:

$$\text{Cov} (X_1, X_2) = \int_{x_1 = -\infty}^{\infty} \int_{x_2 = -\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2)f(x_1, x_2) \, dx_2 \, dx_1$$

$$= \int_{x_1 = -\infty}^{\infty} \frac{x_1 - \mu_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{z_1^2}{2}} \int_{x_2 = -\infty}^{\infty} \frac{x_2 - \mu_2}{\sqrt{2\pi(1 - \rho^2)}\sigma_2} e^{-\frac{(\rho z_1 - z_2)^2}{2(1-\rho^2)}} \, dx_2 \, dx_1$$

$$= \int_{z_1}^{\infty} \int_{z_2}^{\infty} \frac{z_2}{\sqrt{1 - \rho^2}} \phi \left( \frac{\rho z_1 - z_2}{\sqrt{1 - \rho^2}} \right) \sigma_2 \, dz_2 \sigma_1 \, dz_1$$

$$= \sigma_1 \sigma_2 \int_{z_1} \phi(z_1) \rho \, dz_1 = \sigma_1 \sigma_2 \rho = \sigma_{12}$$
Conditional distribution

Interpretation of the formula

\[ f(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1}} e^{-\frac{z_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_2}} e^{-\frac{(\rho z_1 - z_2)^2}{2(1-\rho^2)}} \]

\[ f(x_1, x_2) = f_1(x_1)f_2|_1(x_2|x_1) \]

From \( \frac{(\rho z_1 - z_2)^2}{(1-\rho^2)} = \frac{(\mu_2 + \sigma_2 \rho z_1 - x_2)^2}{\sigma_2^2(1-\rho^2)} \) we conclude:

**Conditional distribution is again normal distribution with**

\[ \mu_{2|1} = \mu_2 + \rho(x_1 - \mu_1)\frac{\sigma_2}{\sigma_1}, \quad \sigma_{2|1} = \sigma_2^2(1 - \rho^2) \]

For bivariate normal distribution: \( \rho = 0 \Rightarrow \text{Independence} \)

In general not correct!
Sum of bivariate normal distributions

Let $X_1, X_2$ be bivariate normal with $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12}$

Then the random variable $Z = X_1 + X_2$ is again normal, with

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\sigma_{12})$$

Proof: For the density of the sum we have

$$f_Z(z) = \int_{-\infty}^{\infty} f(z - x_2, x_2) \, dx_2$$

Get the result again by completion of square (lengthy calculation)

Intuition: Mean and variance of $Z$ obtained from formula for general random variables