50.034 - Introduction to Probability and Statistics

Alexander Binder

Week 11: First lecture on 27th of March Second Lecture on 29th of March

Joint probability distributions

In some experiments, two or more random variables need to be observed simultaneously, to understand their individual behaviors as well as the relationship between them. We have seen this for maximum likelihood estimation already. **Joint probability distributions** help us analyze such situations.

A look back

- In the Thursday Friday lecture the principle of maximum likelihood estimation was derived.
- for a dataset $x_1, \ldots, x_n$ which is modeled by a probability density $f(X_1, \ldots, X_n \mid \theta)$ over n random variables $X_1, \ldots, X_n$, which has a vector of parameters θ, the principle of maximum likelihood estimation states: this parameter vector should be chosen such that the likelihood, that is the density in the data points, is maximized: $\theta_{MLE} = \text{argmax}_\theta f(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid \theta)$
- If one assumes that the random variables are independently and identically distributed, then the density over n variables factors into a product: $f(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid \theta) = f(X_1 = x_1 \mid \theta) \cdot f(X_2 = x_2 \mid \theta) \cdot \ldots \cdot f(X_n = x_n \mid \theta)$. In this case one can make use of 1-dimensional densities $f(X = x \mid \theta)$ that you have seen in the last monday lecture.
- the maximization problem can be solved by setting the gradient to zero, and checking the solution for being a maximum, or, numerically by gradient ascent.
- maximum likelihood is only one way to find parameters. you know the method of moments. You have not seen yet: maximum a posteriori estimation, and fully bayesian methods (ML lecture).

- We needed here probabilities and densities over multiple variables.

We have seen in the lecture on Maximum likelihood estimation a probability distribution over n variables given by a density. Suppose we have n one-dimensional densities $f_i(x_i)$, then

$$f(x_1, \ldots, x_n) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n)$$

defines a density function over n variables (there are many ways to define them!). That is: it is non-negative, and integrates up to 1

$$\begin{aligned}
&f(x_1, \ldots, x_n) \geq 0 \text{ for all } (x_1, \ldots, x_n) \\
&\int_{x_1,\ldots,x_n} f(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1
\end{aligned}$$

The latter holds because

$$\begin{aligned}
&\int_{x_1,\ldots,x_n} f(x_1, \ldots, x_n) dx_1 \cdots dx_n \\
=&\int_{x_1,\ldots,x_n} f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n) dx_1 \cdots dx_n \\
=&\int_{x_1} \int_{x_2} \cdots \int_{x_n} f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n) dx_1 \cdots dx_n \\
=&\int_{x_1} f(x_1) dx_1 \cdot \int_{x_2} f(x_2) dx_2 \cdot \ldots \cdot \int_{x_n} f(x_n) dx_n \\
=&1 \cdot 1 \cdot \ldots \cdot 1 = 1
\end{aligned}$$

This density can be used to define a probability distribution over n variables:

$$\begin{aligned}
&P(X_1 \in [a_1, b_1], X_2 \in [a_2, b_2], \ldots, X_n \in [a_n, b_n]) \\
=&\int_{x_1=a_1}^{x_1=b_1} f(x_1) dx_1 \cdot \int_{x_2=a_2}^{x_2=b_2} f(x_2) dx_2 \cdot \ldots \cdot \int_{x_n=a_n}^{x_n=b_n} f(x_n) dx_n
\end{aligned}$$

Remark[1]

[1] We will see later that this way of construction of a probability of for n variables – using independence – can be either desirable or highly undesirable as it implies independence between variables. If you want to model a joint density for smoking and lung cancer – do you want them to be independent in the model? (Does your dad own a tobacco farm? :))

Joint and Marginal: Two variables, Discrete case

We will consider joint distributions for discrete as well as continuous variables - at first in the case of two variables.

Let X_1 and X_2 be two **discrete random variables** defined on a sample space Ω of an experiment. The **joint probability mass function** $p(x_1, x_2)$ is defined for each pair of members (x_1, x_2) by

$$p(x_1, x_2) = P(X_1 = x_1 \quad and \quad X_2 = x_2) = P(X_1 = x_1, X_2 = x_2)$$

It must be the case that

$$
\begin{aligned}
p(x_1, x_2) &\geq 0 \\
\sum_{x_1} \sum_{x_2} p(x_1, x_2) &= 1
\end{aligned}
$$

The connection between sets and events of joint random variables: What is the meaning of:
$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1 \quad and \quad X_2 = x_2)$?
Probability is defined on sets $S \subset V$ in the space V. Where are these sets S in the definition of $P(X_1 = x_1, X_2 = x_2)$?

Remember that a random variable X_1 maps the space V onto your experiment outcomes. $X_1(v)$ can be x_1 or x_2 or any other value of the possible experiment outcomes.

$P(X_1 = x_1, X_2 = x_2)$ is the probability measure of the set of all $v \in V$ such that $X_1(v) = x_1$ and $X_2(v) = x_2$

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2) = P(S) \\
&\qquad S = \{v \in V \text{ such that } X_1(v) = x_1 \text{ and } X_2(v) = x_2\}
\end{aligned}
$$

Note that the **comma ","** in $P(X_1 = x_1, X_2 = x_2)$ implicitly denotes an **AND** between sets of events.
This establishes the link between the sets used in defining probability and events between multiples variables.

Joint probability over a set A:

Now let A be any set consisting of pairs of (x_1, x_2) values (e.g. $A = \{(x_1, x_2) : x_1, x_2 = 5\}$ or $\{(x_1, x_2) : max(x_1, x_2) \leq 3\}$.

Then the probability $P((X_1, X_2) \in A)$ is obtained by summing the joint pmf over all pairs (x_1, x_2) in A:

$$P((X_1, X_2) \in A) = \sum_{(x_1,x_2)\in A} p(x_1, x_2)$$

The **marginal probability mass function of X_1,** denoted by $p_{X_1}(x_1)$, is given by the following for each possible value of x_1

$$p_{X_1}(x_1) = \sum_{x_2} p(x_1, x_2)$$

Similarly, the **marginal probability mass function of** X_2, denoted by $p_{X_2}(x_2)$, is given by the following for each possible value of x_2

$$p_{X_2}(x_2) = \sum_{x_1} p(x_1, x_2)$$

The marginal probability is the probability distribution of a subset of all variables. The importance of marginal probabilities is three-fold:

1. They provide probabilities defined over a subset of variables (here for one variable only, see below the case for n variables).
2. They are used to define conditional probabilities.
3. They can be used to validate independence - we will see this later.

Joint and Marginal: Two variables, Continuous case

Let X_1 and X_2 be **continuous** random variables. A **joint probability density function** $f(x_1, x_2)$ for these two variables is a function satisfying

$$\begin{aligned} f(x_1, x_2) &\geq 0 \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2 &= 1 \end{aligned}$$

Then for any two-dimensional set A

$$P((X_1, X_2) \in A) = \int_A f(x_1, x_2) dxdy = \int_{(x_1,x_2)\in A} f(x_1, x_2) dx_1 dx_2$$

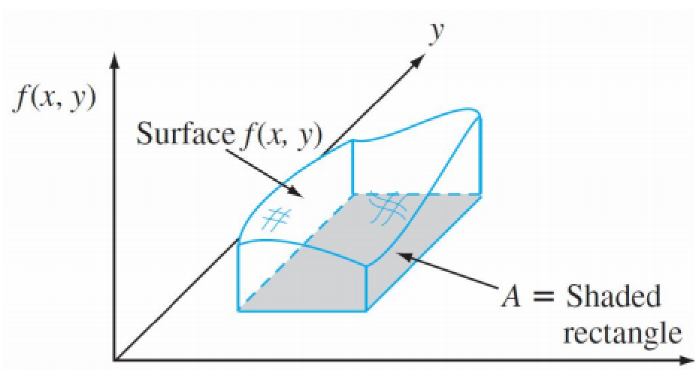
In particular, if A is the two-dimensional rectangle

$[a_1, b_1] \times [a_2, b_2] = \{(x_1, x_2) : a_1 \leq x_1 \leq b_1, a_2 \leq x_2 \leq b_2\}$, then

$$P((X_1, X_2) \in A) = P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) \quad = \int_{x_1=a_1}^{x_1=b_1} \int_{x_2=a_2}^{x_2=b_2} f(x_1, x_2) dx_1 dx_2$$

It helps to compare with the discrete case: **summing is replaced by integration**. Integration is the extension of summation to non-countable spaces like the interval (a, b).

We can think of $f(x_1, x_2)$ as specifying a surface at a height $f(x_1, x_2)$ above the point (x_1, x_2) in a three-dimensional coordinate system. Then $P[(X_1, X_2) \in A]$ is the volume underneath this surface and above the region A, analogous to the area under a curve in the case of a single continuous random variable.



$\mathbb{P}\{(X, Y) \in A\}$ = volume under density surface above A

The **marginal probability density functions** of X_1 and X_2, denoted by $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$, respectively, are given by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$$

If the definition range is not $(-\infty, \infty)$, then replace it by the right interval (a, b).

2 *Variables: Comparison Discrete vs Continuous case*

for probabilities:

$$P((X_1, X_2) \in A) = \sum_{(x_1,x_2)\in A} p(x_1, x_2), \; - p \text{ is pmf}$$
$$P((X_1, X_2) \in A) = \int_{(x_1,x_2)\in A} f(x_1, x_2)\, dx_1 dx_2, \; - f \text{ is pdf}$$

for marginals:

$$p_{X_1}(x_1) = \sum_{x_2} p(x_1, x_2)$$
$$p_{X_1}(x_1) = \int_{x_2 \in \text{ its Definition range}} f(x_1, x_2) dx_2$$
$$p_{X_2}(x_2) = \sum_{x_1} p(x_1, x_2)$$
$$p_{X_2}(x_2) = \int_{x_1 \in \text{ its Definition range}} f(x_1, x_2) dx_1$$

In order to get the marginal in the variable of interest, we must sum/integrate over the variable that we are not interested in.

Joint Distribution: N variables

Now lets consider that we have n variables: $X_1, X_2, \ldots, X_n$. Then the joint and marginal distribution are defined analogously:

$$P((X_1, X_2, \ldots, X_n) \in A) = \sum_{(x_1,x_2,\ldots,x_n)\in A} p(x_1, x_2, \ldots, x_n)$$
$$P((X_1, X_2, \ldots, X_n) \in A) = \int_A f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n$$

In case that the set A is a product of intervals,

$$A = [c_1, d_1] \times [c_2, d_2] \times \ldots [c_n, d_n]$$

then the last integral can be written in a more concrete form:

$$P((X_1, X_2, \ldots, X_n) \in A) = P((X_1, X_2, \ldots, X_n) \in [a_1, b_1] \times [a_2, b_2] \times \ldots [a_n, b_n])$$
$$= \int_{[a_1,b_1]\times[c_2,d_2]\times\ldots[a_n,b_n]} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n$$
$$= \int_{x_1=a_1}^{b_1} \int_{x_2=a_2}^{b_2} \cdots \int_{x_n=a_n}^{b_n} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n$$

For sets A, that do not have such shape, the integral can be tried to be computed using **Integral substitution theorem** for n dimensions[2]

[2] `https://en.wikipedia.org/wiki/Integration_by_substitution`

Marginal Distribution: N variables

The marginal distribution is the distribution of a subset of the set of all variables. The idea is the same: summing away variables. However the set of all possible marginals is now more rich compared to the case of two variables.

For a pmf $P(X_1, X_2, \dots, X_n)$ respectively pdf $f(X_1, X_2, \dots, X_n)$ we can compute marginals with $n-1, n-2, n-3, \dots, 3, 2, 1$ variables.

Example 1: Marginal eliminating X_1:

For example, in the discrete case:

$$P(X_2, \dots, X_n) = \sum_{X_1} P(X_1, X_2, \dots, X_n)$$

and in the continuous case:

$$f(x_2, \dots, x_n) = \int_{x_1 \in Def(X_1)} f(x_1, x_2, \dots, x_n) dx_1$$

where $Def(X_1)$ is the space on which the random variable X_1 is defined. This density can be used to compute a joint probability over the set of variables $(X_2, X_3, X_4, \dots, X_n)$ as done above.

Whether a probability density is a marginal or a joint density – that depends on the set of variables which you consider. $f(x_2, \dots, x_n)$ is a marginal density for the set of variables $(X_1, X_2, \dots, X_n)$, and a joint density for the set of variables $(X_2, \dots, X_n)$.

Example 2: Marginal eliminating X_2:

Another example, in the discrete case, now a marginal without X_2

$$P(X_1, X_3 \dots, X_n) = \sum_{X_2} P(X_1, X_2, \dots, X_n)$$

and in the continuous case:

$$f(x_1, x_3, \dots, x_n) = \int_{x_2 \in Def(X_2)} f(x_1, x_2, \dots, x_n) dx_2$$

This density can be used to compute a joint probability over the set of variables $(X_1, X_3, \dots, X_n)$ as done above.

We can define an marginal probability (discrete case) respectively density (continuous case) for $n-2$ variables, eliminating X_2, X_3:

Example2: Marginal eliminating X_2 and X_3:

$$f(x_1, x_4, \dots, x_n) = \int_{x_2 \in Def(X_2)} \int_{x_3 \in Def(X_3)} f(x_1, x_2, \dots, x_n) dx_2 dx_3$$

This density can be used to compute a joint probability over the set of variables $(X_1, X_4, X_5, \ldots, X_n)$ as done above.

Rule to compute marginal pmfs and pdfs:

The general rule is:

- If one wants to compute the marginal probability P for k variables in the discrete case, then one must sum over the other/remaining $n-k$ variables.
- If one wants to compute the marginal density f for k variables in the continuous case, then one must integrate over the other/remaining $n-k$ variables.

For example consider the marginal density for variables X_2, X_3, X_7 for a set of variables $(X_1, \ldots, X_9)$

$$f(X_2, X_3, X_7) = \int_{x_1} \int_{x_4} \int_{x_5} \int_{x_6} \int_{x_8} \int_{x_9} \cdots \int_{x_n} f(x_1, x_2, \ldots, x_n) dx_1 dx_4 dx_5 dx_6 dx_8 dx_9$$

Question 1

Suppose you have a joint distribution of four variables (X_1, X_2, X_3, X_4). What are the marginals consisting of $3, 2, 1$ variables? How many are them?

Question 2

- How many $n-1$-dimensional marginals we have if we had originally n variables?
- How many 1-dimensional marginals we have if we had originally n variables?
- How many 2-dimensional marginals we have if we had originally n variables?
- How many k-dimensional marginals we have if we had originally n variables?

Joint Cumulative Distribution: Two variables

Assume X_1 and X_2 are two random variables. We define the joint cumulative distribution function as $F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$.

If X_1 and X_2 are continuous random variables, then

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = \int_{y_1=-\infty}^{x_1} \int_{y_2=-\infty}^{x_2} f(y_1, y_2) dy_1 dy_2.$$

This formula is an application of the definition of joint probability for a set A

$$P(X_1 \leq x_1, X_2 \leq x_2) = P(-\infty < X_1 \leq x_1, -\infty < X_2 \leq x_2)$$

using rectangular intervals

$$A = (-\infty, x_1] \times (-\infty, x_2]$$

So there is no need to memorize much here, just apply to $P(-\infty < X_1 \leq x_1, -\infty < X_2 \leq x_2)$.

Relationship between joint cdf and pdf:

$$\frac{\partial^2 F}{\partial x \partial y}(x_1, x_2) = f(x_1, x_2).$$

For example, if $F(x_1, x_2) = x_1 x_2$, then $f(x_1, x_2) = 1$.

Recap on partial derivatives:

$$\frac{\partial^2 F}{\partial x_1 \partial x_2}(x_1, x_2) = \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}F(x_1, x_2)$$

By the Theorem of Schwarz (also called Theorem of Clairaut): if all second derivatives do exist and are continuous, then the order of differentiation does not matter and we have:

$$\frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2}F(x_1, x_2) = \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1}F(x_1, x_2)$$

Joint Cumulative Distribution: N variables

Assume $X_1, X_2, \ldots, X_n$ are random variables. We define the joint cumulative distribution function as $F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$.

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) =$$
$$= \int_{y_1=-\infty}^{x_1} \int_{y_2=-\infty}^{x_2} \cdots \int_{y_n=-\infty}^{x_n} f(y_1, y_2, \ldots, y_n) dy_1 dy_2 \cdots dy_n$$

Same as for 2 Variables, This formula is an application of the definition of joint probability for a set A

$$P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = P(-\infty < X_1 \leq x_1, -\infty < X_2 \leq x_2, \ldots, -\infty < X_n \leq x_n)$$

using rectangular intervals

$$A = (-\infty, x_1] \times (-\infty, x_2] \times \ldots \times (-\infty, x_n]$$

Relationship between joint cdf and pdf:

$$\frac{\partial^n F}{\partial y_1 \partial y_2 \ldots \partial y_n}(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n).$$

For n variables, it is a partial derivative over all n variables.

Expectation of a function of N variables $h(X_1, \ldots, X_n)$

Proposition: Let X_1 and X_2 be jointly distributed random variables with pmf $p(x_1, x_2)$ or pdf $f(x_1, x_2)$ according to whether the variables are discrete or continuous. Then the expected value of a function $h(X_1, X_2)$, denoted by $E[h(X_1, X_2)]$ or $\mu_{h(X_1,X_2)}$, is given by

$$E[h(X_1, X_2)] = \begin{cases} \sum\limits_{x_1}\sum\limits_{x_2} h(x_1, x_2) \cdot p(x_1, x_2), & X_1, X_2 \quad \text{discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x_1, x_2) \cdot f(x_1, x_2) dx_1 dx_2, & X_1, X_2 \quad \text{continuous} \end{cases}$$

This extends directly to n variables. No extra pain:

$$E[h(X_1, \ldots, X_n)] = \int_{x_1}\int_{x_2} \cdots \int_{x_n} h(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

In both last problems we saw integration coming up

Recap: Integral substitution theorem for N variables

As a *recap* (out of class! I write it here for courtesy, because for example wikipedia is for math topics often not a very good choice.) for integral substitution : For one-dimensional integrals (which we do not have here) Integral substitution is

$$\int_{t(a)}^{t(b)} f(x)dx = \int_a^b f(t(y))t'(y)\, dy$$

where $t : [a, b] \subset \mathbb{R}^1 \to \mathbb{R}^1$ is a differentiable one-to-one (bijective) mapping of the interval $[a, b]$ to real numbers.

For sets $A = t(B)$ in high dimensions we have to use:

$$\int_A f(x)dx = \int_{t(B)} f(x)dx = \int_B f(t(y))|det(Dt(y))|\, dy$$

where $t : B \subset \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable one-to-one mapping of the set A in a vector space $\mathbb{R}^n$ onto a part of the vector space $\mathbb{R}^n$. $\mathbb{R}^n$ is the space of n-dimensional vectors composed of real numbers $v = (v_1, \ldots, v_n)$.

$det(M)$ is the determinant of a $n \times n$ matrix. $Dt(y)$ is the $n \times n$ Jacobi-Matrix of all first derivatives of the mapping t.

To understand why the derivative is a matrix, consider this:

$$Dt(y) = \begin{pmatrix} \frac{\partial t_1}{\partial y_1}(y) & \frac{\partial t_2}{\partial y_1}(y) & \cdots & \frac{\partial t_n}{\partial y_1}(y) \\ \frac{\partial t_1}{\partial y_2}(y) & \frac{\partial t_2}{\partial y_2}(y) & \cdots & \frac{\partial t_n}{\partial y_2}(y) \\ \frac{\partial t_1}{\partial y_3}(y) & \frac{\partial t_2}{\partial y_3}(y) & \cdots & \frac{\partial t_n}{\partial y_3}(y) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial t_1}{\partial y_n}(y) & \frac{\partial t_2}{\partial y_n}(y) & \cdots & \frac{\partial t_n}{\partial y_n}(y) \end{pmatrix}$$

t maps B onto a part of the vector space $\mathbb{R}^n$. So $t(y)$ is always an n-dimensional vector $(t_1(y), \ldots, t_n(y))$. Each component $t_i(y)$ takes an n-dimensional vector $y = (y_1, \ldots, y_n)$ as input. Each component $t_i(y)$ has n derivatives - one for each dimension of its input vector $y = (y_1, \ldots, y_n)$. We have a n-element vector of components $t_i(y)$ and each of its components has n derivatives – this gives an $n \times n$ matrix. The determinant is calculated for this matrix.

You can also define this matrix as its transpose.

Worked-out examples related to week's first lecture

1. A large insurance agency serves a number of customers who have purchased both home insurance policy and car insurance policy from the agency. For each type of policy, a deductible amount[3] must be specified. For a car policy, the choices are \$100 and \$250, whereas for a home policy, the choices are \$0, \$100, and \$200. Suppose an individual with both types of policy is selected at random from the agency's files. Let X_1 = the deductible amount on the car policy and X_2 = the deductible amount on the home policy. The joint probability mass function is given in the following table.

[3] A deductible amount in an insurance policy is the amount that must be paid by the policyholder in the event of a claim, before receiving any reimbursement from the insurance company.

		x_2		
$p(x_1, x_2)$		0	100	200
x_1	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

What is $P(X_2 \geq 100)$?

2. In the earlier example about the large insurance agency, with the following joint probability table, what are the marginal probability mass functions of X_1 and X_2?

$p(x_1, x_2)$		x_2: 0	100	200
x_1	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

3. You are part of a start-up company. Your company responds to customer concerns through two channels: telephone helpline and online chat. Since there are very few of you in the company, often the same people need to attend to both channels. Let X_1 = the proportion of time that day the telephone helpline is in use (at least one customer being served or waiting to be served) and X_2 = the proportion of time that the online chat channel is in use. Suppose the joint probability density function of (X_1, X_2) is given by:

$$f(x_1, x_2) = \begin{cases} \frac{6}{5}(x_1 + x_2^2) & 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

What is the probability that neither channel is busy more than one-quarter of the time?

4. Using the joint probability density function of (X_1, X_2) (given below) of the earlier example involving a start-up company operating a telephone helpline, as well as an online chat channel,

$$f(x_1, x_2) = \begin{cases} \frac{6}{5}(x_1 + x_2^2) & 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

   (a) What are the marginal probability density functions of X_1 and X_2?

   (b) What is the probability that the online chat channel is used between one-quarter and three-quarter of the time?

5. Tire pressure is a critical safety concern for airplanes during landing and takeoff.Let us assume, each rear tire on an airplane is supposed to be filled to a pressure of 40 pounds per square inch (psi). Let X_1 denote the actual air pressure for the right tire and X_2 denote the actual air pressure for the left tire. Suppose that X_1 and X_2 are random variables with the joint density function

$$f(x_1, x_2) = \begin{cases} k(x_1^2 + x_2^2), & 30 \leq x_1 < 50, 30 \leq x_2 < 50 \\ 0, & \text{elsewhere} \end{cases}$$

Find the probability that both tires are under-filled.

6. In a cohort I will be teaching next term, I am thinking of giving out different types of chocolates. There will be 3 types of chocolates; let us call them A, B, and C. I will buy some amounts of these chocolates and randomly mix them in a box to bring to class. The total weight of chocolates in the box is 1 lb and the joint distribution of the weight X_1 of A and the weight X_2 of B in the 1 lb box is:

$$f(x_1, x_2) = \begin{cases} 24x_1x_2 & 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, x_1 + x_2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

If 1 lb of A costs \$1.00, 1 lb of B costs \$1.50 and 1 lb of C costs \$ 0.50, what is the expected total cost of the 1 lb of mixed chocolates in the box?

7. Suppose you measure wind speed vectors. The vectors are parallel to the surface, so we assume that they are two dimensional, that is $v = (x, y)$. You can compute the magnitude of the speed as $m = \sqrt{x^2 + y^2}$ – this is the euclidean length in a two dimensional space.

   The density is given as

$$f(x,y) = \begin{cases} k\exp(-(x^2+y^2)) & \text{if } x^2+y^2 < 3 \cdot 3 = 9 \\ 0 & \text{else} \end{cases}$$

   How k has to be chosen, so that it becomes a probability density? What is the probability to observe a magnitude of the windspeed of larger than $1.5m/s$

8. .