

Transformation of Probability Densities

This Wikibook shows how to transform the probability density of a **continuous random variable** in both the one-dimensional and multidimensional case. In other words, it shows how to calculate the distribution of a function of continuous random variables. The first section formulates the general problem and provides its solution. However, this general solution is often quite hard to evaluate and simplifications are possible in special cases, e.g., if the random vector is one-dimensional or if the components of the random vector are independent. Formulas for these special cases are derived in the subsequent sections. This Wikibook also aims to provide an overview of different equations used in this field and show their connection.

1 General Problem and Solution (n-to-m mapping)

Let $\vec{X} = (X_1, \dots, X_n)$ be a **random vector** with the **probability density function**, pdf, $\varrho_{\vec{X}}(x_1, \dots, x_n)$ and let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a (**Borel-measurable**) function. We are looking for the probability density function $\varrho_{\vec{Y}}$ of $\vec{Y} := \vec{f}(\vec{X})$. First, we need to remember the definition of the **cumulative distribution function**, cdf, $F_{\vec{Y}}(\vec{y})$ of a random vector: It measures the probability that each component of \vec{Y} takes a value smaller than the corresponding component of \vec{y} . We will use a short-hand notation and say that two vectors are "less or equal" (\leq) if all of their components are.

$$F_{\vec{Y}}(\vec{y}) = P(\vec{Y} \leq \vec{y}) = P(\vec{f}(\vec{X}) \leq \vec{y}) \quad (1)$$

The wanted density $\varrho_{\vec{Y}}(\vec{y})$ is then obtained by differentiating $F_{\vec{Y}}(\vec{y})$:

$$\varrho_{\vec{Y}}(\vec{y}) = \frac{\partial}{\partial y_1} \cdots \frac{\partial}{\partial y_m} F_{\vec{Y}}(\vec{y}) \quad (2)$$

Thus, the general solution can be expressed as the m'th derivative of an n-dimensional integral:

$$\boxed{\varrho_{\vec{Y}}(\vec{y}) = \frac{\partial}{\partial y_1} \cdots \frac{\partial}{\partial y_m} \int_{\{\vec{x} \in \mathbb{R}^n | \vec{f}(\vec{x}) \leq \vec{y}\}} \varrho_{\vec{X}}(\vec{x}) d^n x} \quad (3)$$

The following sections will provide simplifications in special cases.

2 Function of a Random Variable (n=1, m=1)

If n=1 and m=1, X is a continuously distributed **random variable** with the density ϱ_X and $f : \mathbb{R} \rightarrow \mathbb{R}$ is a Borel-measurable function. Then also $Y := f(X)$ is continuously distributed and we are looking for the density $\varrho_Y(y)$. In the following, f should always be at least differentiable.

Let us first note that there may be values that are never reached by f, e.g. $y < 0$ if $f(x) = x^2$. For all those y, necessarily $\varrho_Y(y) = 0$.

$$\varrho_Y(y) = \begin{cases} 0, & \text{if } y \notin f(\mathbb{R}) \\ ?, & \text{if } y \in f(\mathbb{R}) \end{cases}$$

Following equations (1) and (2), we obtain

$$\varrho_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \leq y) = \frac{d}{dy} P(f(X) \leq y) \quad (4)$$

We will now rearrange this expression in various ways.

2.1 Deviation using cumulated distribution function

F_X

At first, we limit ourselves to f those derivative is never 0 (thus, f is a **diffeomorphism**). Then, the inverse map f^{-1} exists and f is either monotonically increasing or monotonically decreasing.

If f is monotonically increasing, then $x \leq f^{-1}(y) \Leftrightarrow f(x) \leq f(f^{-1}(y)) = y$ and $f' > 0$. Therefore:

$$\begin{aligned} \varrho_Y(y) &= \frac{d}{dy} P(f(X) \leq y) = \frac{d}{dy} P(X \leq f^{-1}(y)) \\ &= \frac{d}{dy} F_X(f^{-1}(y)) = \varrho_X(f^{-1}(y)) \frac{df^{-1}(y)}{dy} \end{aligned}$$

If f is monotonically decreasing, then $x \leq f^{-1}(y) \Leftrightarrow f(x) \geq f(f^{-1}(y)) = y$ and $f' < 0$. Therefore:

$$\begin{aligned} \varrho_Y(y) &= \frac{d}{dy} P(f(X) \leq y) = \frac{d}{dy} P(X \geq f^{-1}(y)) \\ &= \frac{d}{dy} (1 - F_X(f^{-1}(y))) = -\varrho_X(f^{-1}(y)) \frac{df^{-1}(y)}{dy} \end{aligned}$$

This can be summarized as:

$$\varrho_Y(y) = \begin{cases} 0, & \text{if } y \notin f(\mathbb{R}) \\ \varrho_X(f^{-1}(y)) \cdot \left| \frac{df^{-1}(y)}{dy} \right|, & \text{if } y \in f(\mathbb{R}) \end{cases} \quad (5)$$

If now the derivative $f'(x_i) = 0$ does vanish at some positions x_i , $i = 1, \dots, N$, then we split the definition space of f using those position into $N + 1$ disjoint intervals I_j . Equation (5) holds for the functions f_{I_j} limited in their definition space to those intervals I_j . We have

$$P(f(X) \leq y) = \sum_{j=1}^{N+1} P(f_{I_j}(X) \leq y)$$

$$\varrho_Y(y) = \frac{d}{dy} P(f(X) \leq y) = \sum_{j=1}^{N+1} \varrho_X(f_{I_j}^{-1}(y)) \cdot \left| \frac{df_{I_j}^{-1}(y)}{dy} \right|$$

With the convention that the sum over 0 addends is 0 and using the inverse function theorem, it is possible to write this in a more compact form (read as: sum over all x, where f(x)=y):

$$\mathbb{R} \rightarrow \mathbb{R} \text{ mapping}$$

$$\varrho_Y(y) = \sum_{x, f(x)=y} \frac{\varrho_X(x)}{|f'(x)|}$$

(6)

2.2 Deviation using Integral Substitution

In this section we consider a different deviation.

The probability in (4) is the integral over the probability density. Again in the case of monotonically increasing f, we have:

$$\int_{-\infty}^y \varrho_Y(u) du = P(Y \leq y) = P(f(X) \leq y) = P(X \leq f^{-1}(y))$$

$$= \int_{-\infty}^{f^{-1}(y)} \varrho_X(x) dx$$

Now we substitute u = f(x) in the integral on the right hand site, i.e. x = f⁻¹(u) and $\frac{du}{dx} = f'(x)$. The integral limits are then from -∞ to y and in the rule “ $dx = \frac{dx}{du} du$ ” we have $\frac{dx}{du} = \frac{df^{-1}(u)}{du}$ due to the inverse function theorem. Consequentially:

$$\int_{-\infty}^y \varrho_Y(u) du = \int_{-\infty}^y \varrho_X(f^{-1}(u)) \frac{df^{-1}(u)}{du} du$$

Taking the derivative of both sides with respect to y, we get:

$$\varrho_Y(y) = \varrho_X(f^{-1}(y)) \frac{df^{-1}(y)}{dy}$$

Following the same argument as in the last section, we can again derive equation (6).

This rule often misleads physics books to present the following viewpoint, which might be easier to remember, but is not mathematically sound: If you multiply the probability density $\varrho_X(x)$ with the “infinitesimal length” dx , then you will get the probability $\varrho_X(x) dx$ for X to lie in the interval [x, x+dx]. Changing to new coordinates y, you will get by substitution:

$$\varrho_X(x) dx = \underbrace{\varrho_X(x(y)) \frac{dx}{dy}}_{\varrho_Y(y)} dy$$

2.3 Deviation using the Delta Distribution

In this section we consider another different deviation, often used in physics. We start again with equation (4) and write this as an integral:

$$\begin{aligned}
 \varrho_Y(y) &= \frac{d}{dy} P(f(X) \leq y) \\
 &= \frac{d}{dy} \int_{\{x \in \mathbb{R} | f(x) \leq y\}} \varrho_X(x) dx \\
 &= \frac{d}{dy} \int_{\mathbb{R}} \Theta(y - f(x)) \varrho_X(x) dx \\
 &= \int_{\mathbb{R}} \frac{d}{dy} \Theta(y - f(x)) \varrho_X(x) dx \\
 &= \int_{\mathbb{R}} \delta(y - f(x)) \varrho_X(x) dx
 \end{aligned}$$

The intuitive interpretation of the last expression is: one integrates over all possible x-values and uses the **delta “function”** to pick all positions where $y = f(x)$. This formula is often found in physics books, possibly written as expectation value, $\langle \dots \rangle$:

$$\boxed{\begin{array}{l} \mathbb{R} \rightarrow \mathbb{R} \text{ mapping (using Dirac Delta Distribution)} \\ \varrho_Y(y) = \int_{\mathbb{R}} \delta(y - f(x)) \varrho_X(x) dx = \langle \delta(y - f(x)) \rangle . \end{array}} \quad (7)$$

We can see this formula is equivalent to equation (6) using the following identity

$$\int_{\mathbb{R}} \delta(h(x)) g(x) dx = \sum_{x, h(x)=0} \frac{g(x)}{|h'(x)|}$$

2.4 Example

- Let us consider the following specific example: let $\varrho_X(x) = \frac{\exp[-0.5x^2]}{\sqrt{2\pi}}$ and $f(x) = x^2$. We choose to use equation (6) (equations (5) and (7) lead to the same result). We calculate the derivative $f'(x) = 2x$ and find all x for which $f(x)=y$, which are $-\sqrt{y}$ and $+\sqrt{y}$ if $y>0$ and none otherwise. For $y>0$ we have:

$$\begin{aligned}
 \varrho_Y(y) &= \sum_{x, f(x)=y} \frac{\varrho_X(x)}{|f'(x)|} = \frac{\varrho_X(-\sqrt{y})}{|f'(-\sqrt{y})|} + \frac{\varrho_X(+\sqrt{y})}{|f'(+\sqrt{y})|} = \frac{\exp[-0.5y]}{\sqrt{2\pi} 2\sqrt{y}} + \frac{\exp[-0.5y]}{\sqrt{2\pi} 2\sqrt{y}} \\
 &= \frac{\exp[-0.5y]}{\sqrt{2\pi} y}
 \end{aligned}$$

Since f never reaches negative values, the sum remains 0 for $y<0$ and we finally obtain:

$$\varrho_Y(y) = \begin{cases} 0, & \text{if } y \leq 0 \\ \frac{\exp[-0.5y]}{\sqrt{2\pi} y}, & \text{if } y > 0 \end{cases}$$

This example is illustrated in figure 1.

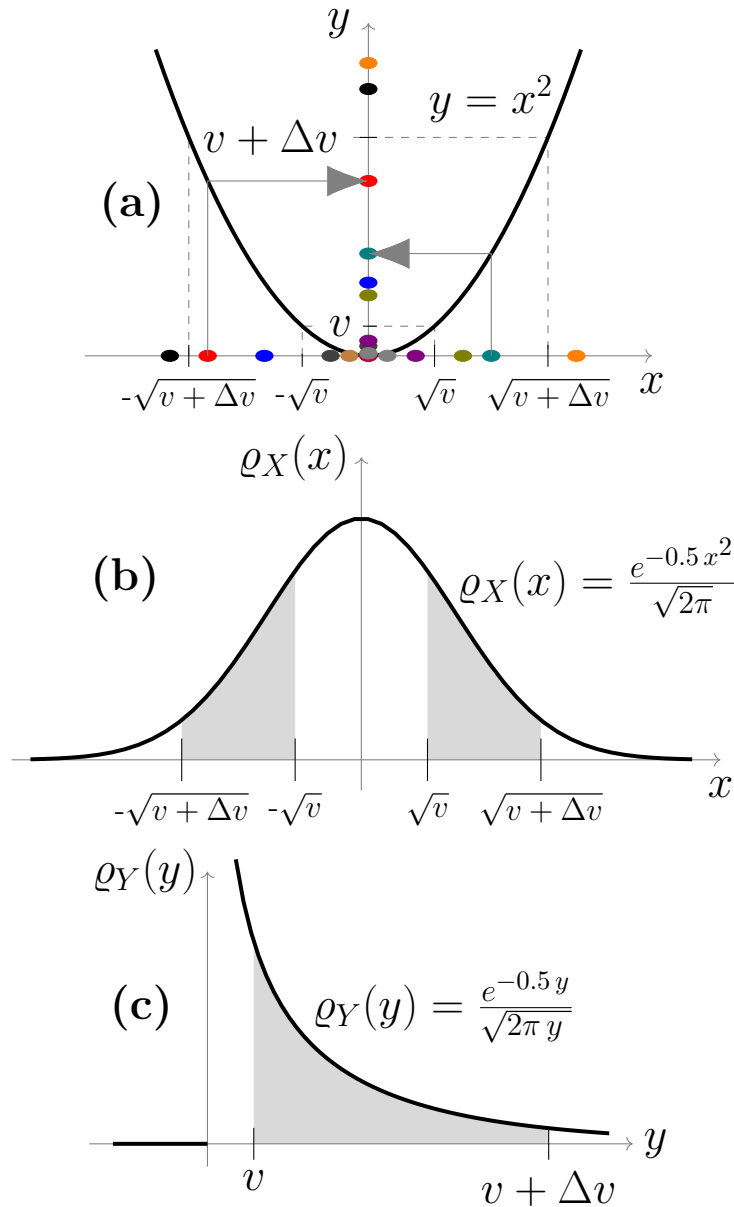


Figure 1: Random numbers are generated according to the standard normal distribution $X \sim N(0,1)$. They are shown on the x -axis of figure (a). Many of them are around 0. Each of those numbers is mapped according to $y = x^2$, which is shown with grey arrows for two example points. For many generated realisations of X , a histogram on the y -axis will converge towards the wanted probability density function ϱ_Y shown in (c). In order to analytically derive this function, we start by observing that in order for Y to be between any v and $v + \Delta v$, X must have been between either $-\sqrt{v + \Delta v}$ and $-\sqrt{v}$, or, between \sqrt{v} and $\sqrt{v + \Delta v}$. Those intervals are marked in figures (b) and (c). Because the probability is equal to the area under the probability density function, we can determine ϱ_Y from the condition that the grey shaded area in (c) must be equal to the sum of the areas in (b). The areas are calculated using integrals and it is useful to take the limit $\Delta v \rightarrow 0$ in order to get the formula noted in (c).

- Another example is the **inverse transformation method**. Suppose a computer generates random numbers X with a uniform distribution on $[0, 1]$, i.e.

$$\varrho_X(x) \equiv \begin{cases} 1, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{else.} \end{cases}$$

If we want to obtain random numbers according to a distribution with the pdf ϱ_Z , we choose f as the inverse function of the cdf of Z , i.e. $Y = f(X) = F_Z^{-1}(X)$. We can now show that Y will have the same distribution as the wanted Z , $\varrho_Y = \varrho_Z$ by using equation (5) and the fact that $f^{-1} = F_Z$:

$$\begin{aligned} \varrho_Y(y) &= \begin{cases} 0, & \text{if } y \notin F_Z^{-1}(\mathbb{R}) \\ \varrho_X(F_Z(y)) \cdot \left| \frac{dF_Z(y)}{dy} \right|, & \text{if } y \in F_Z^{-1}(\mathbb{R}) \end{cases} . \\ &= 1 \cdot \left| \frac{dF_Z(y)}{dy} \right| = \varrho_Z(y) \end{aligned}$$

An example is illustrated in figure 2.

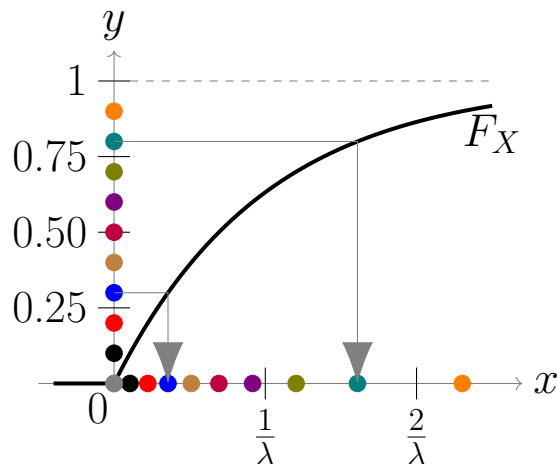


Figure 2: Random numbers y_i are generated from a uniform distribution between 0 and 1, i.e. $Y \sim U(0, 1)$. They are sketched as colored points on the y -axis. Each of the points is mapped according to $x=F^{-1}(y)$, which is shown with gray arrows for two example points. In this example, we have used an exponential distribution. Hence, for $x \geq 0$, the probability density is $\varrho_X(x) = \lambda e^{-\lambda x}$ and the cumulated distribution function is $F(x) = 1 - e^{-\lambda x}$. Therefore, $x = F^{-1}(y) = -\frac{\ln(1-y)}{\lambda}$. We can see that using this method, many points end up close to 0 and only few points end up having high x -values - just as it is expected for an exponential distribution.

3 Mapping of a Random Vector to a Random Variable ($n > 1$, $m = 1$)

We will now investigate the case when a random vector X with known density $\varrho_{\vec{X}}$ is mapped to (scalar) random variable Y and calculate the new density $\varrho_Y(y)$.

According to (3), we find:

$$\varrho_Y(y) = \frac{d}{dy} \int_{\{\vec{x} \in \mathbb{R}^n | f(\vec{x}) \leq y\}} \varrho_{\vec{X}}(\vec{x}) d^n x \quad (8)$$

The direct evaluation of this equation is sometimes the easiest way, e.g., if there is a known formula for the area or volume presented by the integral. Otherwise one needs to solve a parameter-dependent multiple integral.

If the components of the random vector \vec{X} are independent, then the probability density factorizes:

$$\varrho_{\vec{X}}(x_1, \dots, x_n) = \varrho_{X_1}(x_1) \cdot \dots \cdot \varrho_{X_n}(x_n)$$

In this case the delta function may provide a fast tool for the evaluation:

$$\begin{aligned} \varrho_Y(y) &= \int_{\mathbb{R}^n} \varrho_{\vec{X}}(x_1, \dots, x_n) \delta(y - f(\vec{x})) dx_1 \dots dx_n \\ &= \int_{\mathbb{R}} \varrho_{X_n}(x_n) \dots \int_{\mathbb{R}} \varrho_{X_1}(x_1) \delta(y - f(\vec{x})) dx_1 \dots dx_n \end{aligned} \quad (9)$$

If one wants to avoid calculations with the delta function, it is of course possible to evaluate the innermost integral $\int dx_1$, provided that the components are independent:

$$\varrho_Y(y) = \int_{\mathbb{R}^{n-1}} \sum_{x_1, f(\vec{x})=y} \frac{\varrho_{\vec{X}}(\vec{x})}{\left| \frac{\partial f(\vec{x})}{\partial x_1} \right|} dx_2 \dots dx_n$$

3.1 Examples

- Let $Y = f(\vec{X}) = X_1 + X_2$ with the independent, continuous random variable X_1 and X_2 . According to equation (9), we have:

$$\begin{aligned} \varrho_Y(y) &= \int_{\mathbb{R}} \varrho_{X_2}(x_2) \int_{\mathbb{R}} \varrho_{X_1}(x_1) \delta(y - x_2 - x_1) dx_1 dx_2 \\ &= \int_{\mathbb{R}} \varrho_{X_2}(x_2) \varrho_{X_1}(y - x_2) dx_2 \end{aligned}$$

If one uses the sum formula instead, the sum runs over all x_1 , for which $f(\vec{x}) = x_1 + x_2 = y$, i.e. where $x_1 = y - x_2$. The derivative is $\frac{\partial(x_1+x_2)}{\partial x_1} = 1$, so that one also obtains the equation $\varrho_Y(y) = \int_{\mathbb{R}} \varrho_{X_2}(x_2) \varrho_{X_1}(y-x_2) dx_2$.

Integrating over x_2 first leads to the following, equivalent expression:

$$\varrho_Y(y) = \int_{\mathbb{R}} \varrho_{X_1}(x_1) \varrho_{X_2}(y - x_1) dx_1$$

- If $Y = X_1 - X_2$ with independent X_1 and X_2 , then

$$\varrho_Y(y) = \int_{\mathbb{R}} \varrho_{X_1}(x_1) \varrho_{X_2}(x_1 - y) dx_1.$$

- If $Y = X_1 \cdot X_2$ with independent X_1 and X_2 , then

$$\varrho_Y(y) = \int_{\mathbb{R}} \varrho_{X_1}(x_1) \varrho_{X_2}\left(\frac{y}{x_1}\right) \frac{1}{|x_1|} dx_1.$$
- If $Y = \frac{X_1}{X_2}$ with independent X_1 and X_2 , then

$$\varrho_Y(y) = \int_{\mathbb{R}} \varrho_{X_1}(x_2 \cdot y) \varrho_{X_2}(x_2) |x_2| dx_2.$$
- Given the independent random variables X_1 and X_2 with the density

$$\varrho_{\vec{X}}(x_1, x_2) = \begin{cases} 1/\pi, & \text{if } x_1^2 + x_2^2 \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Let $Y := \sqrt{X_1^2 + X_2^2}$. According to equation (8), we need to solve:

$$\varrho_Y(y) = \frac{d}{dy} \int_{\{\vec{x} \in \mathbb{R}^n \mid \sqrt{x_1^2 + x_2^2} \leq y\}} \frac{1}{\pi} dx_1 dx_2$$

The last integral is over a circle with radius $y \leq 1$, hence with the area πy^2 . This simplifies the calculation:

$$0 \leq y \leq 1 \Rightarrow \varrho_Y(y) = \frac{1}{\pi} \frac{d}{dy}(\pi y^2) = \frac{2\pi y}{\pi} = 2y.$$

If $y < 0$, we integrate over an empty set, which gives 0. If $y > 1$, $\varrho_{\vec{X}} = 0$. Therefore, the final result is:

$$\varrho_Y(y) = \begin{cases} 2y, & \text{if } 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

This example is illustrated in the figure 3.

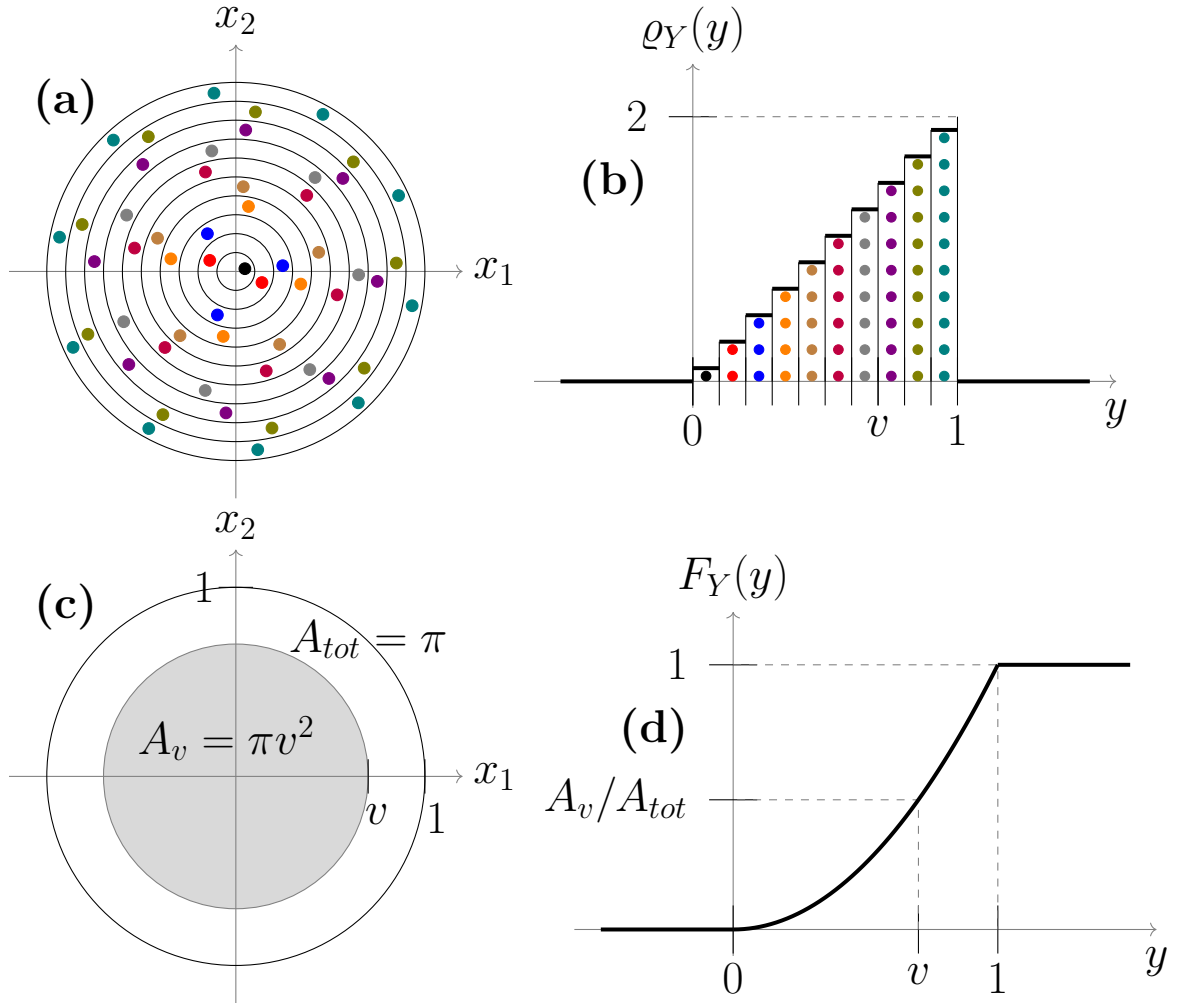


Figure 3: Figure (a) shows a sketch of random sample points from a uniform distribution in a circle of radius 1, subdivided into rings. In figure (b), we count how many of those points fell into each of the rings with the same width Δv . Since the area of the rings increases linearly with the radius, one can expect more points for larger radii. For $\Delta v \rightarrow 0$, the normalized histogram in (b) will converge to the wanted probability density function ρ_Y . In order to calculate ρ_Y analytically, we first derive the cumulated distribution function F_Y , plotted in figure (d). $F_Y(y)$ is the probability to find a point inside the circle of radius v (shown in grey in figure (c)). For v between 0 and 1, we find $F_Y(v) = \frac{A_v}{A_{tot}} = \frac{\pi v^2}{\pi 1^2} = v^2$. The slope of F_Y is the wanted probability density function $\rho_Y(y) = \frac{dF_Y(y)}{dy} = 2y$, in agreement with figure (b).

4 Invertible Transformation of a Random Vector (n=m)

Let $\vec{X} = (X_1, \dots, X_n)$ be a random vector with the density $\varrho_{\vec{X}}(x_1, \dots, x_n)$ and let $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a diffeomorphism. For the density $\varrho_{\vec{Y}}$ of $\vec{Y} := \vec{f}(\vec{X})$ we have:

$$\int_G \varrho_{\vec{X}}(\vec{x}) d^n x = \int_{f(G)} \varrho_{\vec{X}}(f^{-1}(\vec{y})) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right| d^n y$$

and therefore

$$\boxed{\begin{array}{c} \mathbb{R}^n \rightarrow \mathbb{R}^n \text{ mapping} \\ \varrho_{\vec{Y}}(\vec{y}) = \varrho_{\vec{X}}(f^{-1}(\vec{y})) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|, \end{array}} \quad (10)$$

where $\Phi_{f^{-1}} = \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|$ is the **Jacobian determinant** of f^{-1} . Note that $\Phi_{f^{-1}} = (\Phi_f)^{-1}$. In the one-dimensional case (n=1), equation (10) coincides with equation (5).

4.0.1 Examples

- Given the random vector \vec{X} and the invertible matrix A and a vector \vec{b} , let $\vec{Y} = A\vec{X}^T + \vec{b}$. Then $\varrho_{\vec{Y}}(\vec{y}) = \varrho_{\vec{X}}(A^{-1}(\vec{y} - \vec{b})) |\det A^{-1}|$. Also, $\det A^{-1} = 1/\det A$.
- Given the independent random variables X_1 and X_2 , we introduce **polar coordinates** $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = \text{atan2}(X_2, X_1)$. The inverse map is $X_1 = Y_1 \cos Y_2$ and $X_2 = Y_1 \sin Y_2$. Due to Jacobian determinant y_1 , the wanted density is $\varrho_{\vec{Y}}(y_1, y_2) = y_1 \varrho_{\vec{X}}(y_1 \cos y_2, y_1 \sin y_2)$.

5 Possible simplifications for multidimensional mappings ($n > 1$, $m > 1$)

Even if none of the above special cases apply, simplifications can still be possible. Some of them are listed below:

5.1 Independent Target-Components

If one knows beforehand that the components of Y_i will be independent, i.e.

$$\varrho_{\vec{Y}}(y_1, \dots, y_n) = \varrho_{Y_1}(y_1) \cdot \dots \cdot \varrho_{Y_n}(y_n),$$

then the density ϱ_{Y_i} of each component $Y_i = f_i(\vec{X})$ can be calculated like in the above section [Mapping of a Random Vector to a Random Variable](#).

5.1.1 Example

Given the random vector $\vec{X} = (X_1, X_2, X_3, X_4)$ with independent components. Let $f : \mathbb{R}^4 \rightarrow \mathbb{R}^2$, $\vec{Y} = \vec{f}(\vec{X}) := (X_1 + X_2, X_3 + X_4)$. Obviously, the components $Y_1 = X_1 + X_2$ and $Y_2 = X_3 + X_4$ are independent and therefore:

$$\varrho_{Y_1}(y_1) = \int_{\mathbb{R}} \varrho_{X_1}(y_1 - x_2) \varrho_{X_2}(x_2) dx_2 \text{ and}$$

$$\varrho_{Y_2}(y_2) = \int_{\mathbb{R}} \varrho_{X_3}(y_2 - x_4) \varrho_{X_4}(x_4) dx_4.$$

Note that the components of \vec{Y} can be independent even if the components of \vec{X} are not.

5.2 Split Integral Region

Sometimes it is useful to split the integral region in equation (3) into parts that can be evaluate separately. This can be made explicit by rewriting (3) with delta functions:

$$\varrho_{\vec{Y}}(\vec{y}) = \int_{\mathbb{R}^n} \varrho_{\vec{X}}(\vec{x}) \delta(y_1 - f(\vec{x})_1) \cdot \dots \cdot \delta(y_m - f(\vec{x})_m) d^n x$$

and then use the identity $\delta(x - x_0) = \int_{\mathbb{R}} \delta(x - \xi) \delta(\xi - x_0) d\xi$.

5.2.1 Example

To illustrate the idea, we use a simple $\mathbb{R}^3 \rightarrow \mathbb{R}$ example: Let $Y = X_1^2 + X_2^2 + X_3$ with

$$\varrho_{\vec{X}}(x_1, x_2, x_3) = \begin{cases} \frac{e^{-x_3}}{\pi}, & \text{if } x_1^2 + x_2^2 \leq 1 \wedge x_3 \geq 0 \\ 0, & \text{else} \end{cases}$$

The parametrisation of the region where at the same time $x_1^2 + x_2^2 + x_3 \leq y$ and $x_1^2 + x_2^2 \leq 1$ and $x_3 \geq 0$ may not be obvious, so we use the two above formulas:

$$\begin{aligned}
\varrho_Y(y) &= \int_{\mathbb{R}^3} \varrho_{\vec{X}}(\vec{x}) \delta(y - f(\vec{x})) dx_1 dx_2 dx_3 \\
&= \int_0^\infty \iint_{x_1^2 + x_2^2 \leq 1} \frac{e^{-x_3}}{\pi} \delta(y - x_1^2 - x_2^2 - x_3) dx_1 dx_2 dx_3 \\
&= \int_0^\infty \iint_{x_1^2 + x_2^2 \leq 1} \frac{e^{-x_3}}{\pi} \int_{\mathbb{R}} \delta(\xi - x_1^2 - x_2^2) \delta(y - x_3 - \xi) d\xi dx_1 dx_2 dx_3 \\
&= \int_0^\infty \int_{\mathbb{R}} \left[\iint_{x_1^2 + x_2^2 \leq 1} \frac{e^{-x_3}}{\pi} \delta(\xi - x_1^2 - x_2^2) dx_1 dx_2 \right] \delta(y - x_3 - \xi) d\xi dx_3
\end{aligned}$$

Now, we have split the integral such that the expression in brackets can be evaluated separately, because the region depends on x_1 and x_2 only and may contain x_3 only as parameter.

$$\iint_{x_1^2 + x_2^2 \leq 1} \frac{e^{-x_3}}{\pi} \delta(\xi - x_1^2 - x_2^2) dx_1 dx_2 = \begin{cases} e^{-x_3}, & \text{if } 0 \leq \xi \leq 1 \\ 0, & \text{else} \end{cases}$$

Therefore:

$$\varrho_Y(y) = \int_0^\infty \int_0^1 e^{-x_3} \delta(y - x_3 - \xi) d\xi dx_3 = \int_{\max(0, y-1)}^y e^{-x_3} dx_3 = \begin{cases} e^{1-y} - e^{-y}, & \text{if } y > 1 \\ 1 - e^{-y}, & \text{if } 0 \leq y \leq 1 \\ 0, & \text{if } y \leq 0 \end{cases}$$

5.3 Helper Coordinates

If f is injective, then it can be easier to introduce additional helper coordinates Y_{m+1} to Y_n , then do the $\mathbb{R}^n \rightarrow \mathbb{R}^n$ transformation from section [Invertible Transformation of a Random Vector](#) and finally integrate out all helper coordinates of the so-obtained density.

5.3.1 Example

Given the random vector $\vec{X} = (X_1, X_2, X_3)$ with the density $\varrho_{\vec{X}}(\vec{x})$ and the following mapping:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

Now we introduce the helper coordinate $Y_3 = X_3$, which results in the transformation matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

with the corresponding pdf $\varrho_{\vec{X}}(A^{-1}\vec{y}) |\det A^{-1}|$. Thus, we finally obtain

$$\varrho_{\vec{Y}}(y_1, y_2) = \int_{\mathbb{R}} \varrho_{\vec{X}} \left(A^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right) |\det A^{-1}| dy_3 .$$

Remark: If the joint pdf $\varrho_{\vec{Y}}(y_1, y_2)$, i.e. the [conditional distribution](#), is not of interest and one is only interested in the [marginal distribution](#)

with $\varrho_{Y_1}(y_1) = \int_{\mathbb{R}} \varrho_{\vec{Y}}(y_1, y_2) dy_2$, then it is possible to calculate the density as described in section [Mapping of a Random Vector to a Random Variable](#) for the mapping $Y_1 = 1 X_1 + 2 X_2 + 3 X_3$ (and likewise for $Y_2 = 4 X_1 + 5 X_2 + 6 X_3$).

6 Real-World Applications

In order to show some possible applications, we present the following questions, which can be answered using the techniques outlined in this Wiki-book. In principle, the answers could also be approximated using a numerical random number simulation: generate several realizations of \vec{X} , calculate $\vec{Y} = f(\vec{X})$ and make a histogram of the results. However, many such random numbers are needed for reasonable results, especially for higher-dimensional random vectors. Gladly, we can always calculate the resulting distribution analytically using the above formulas.

6.1 Statistical Physics

- Suppose atoms in a laser are moving with normally distributed velocities V_x , $\varrho_{V_x}(v_x) = \frac{\exp[-v_x^2/2\sigma^2]}{\sqrt{2\pi\sigma^2}}$, $\sigma^2 = k_B T/m$. Due to the Doppler effect, light emitted with frequency f_0 by an atom moving with v_x will be detected as $f \approx f_0 (1 + v_x / c)$. Hence, f is a function of V_x . What does the detected spectrum, ϱ_f , look like? (Answer: Gaussian around f_0 .)
- Suppose the velocity components of an ideal gas (V_x, V_y, V_z) are identically, independently normally distributed as in the last example. What is the probability density ϱ_V of $V = \sqrt{V_x^2 + V_y^2 + V_z^2}$? (The answer is known as Maxwell-Boltzmann distribution.)

6.2 Quantifying Uncertainty of Derived Properties

- Suppose we do not know the exact value of X and Y , but we can assign a probability distribution to each of them. What is the distribution of the derived property $Z = X^2 / Y$ and what is the mean value and standard deviation of Z ? (To tackle such problems, linearisation around the mean values are sometimes used and both X and Y are assumed to be normally distributed. However, we are not limited to such restrictions.)
- Suppose we consider the value of one gram gold, silver and platinum in one year from now as independent random variables G, S and P , respectively. Box A contains 1 gram gold, 2 gram silver and 3 gram platinum.

Box B contains 4, 5 and 6 gram, respectively. Thus, $\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} G \\ S \\ P \end{pmatrix}$.

What is the value of the contents in box A (or box B) in one year from now? (The answer is given in an example above.) Note that A and B are correlated.

Note that the above examples assume the distribution of \vec{X} to be known. If it is unknown, or if the calculation is based on only a few data points, methods from mathematical statistics are a better choice to quantify uncertainty.

6.3 Generation of Correlated Random Numbers

Correlated random numbers can be obtained by first generating a vector of uncorrelated random numbers and then applying a function on them.

- In order to obtain random numbers with covariance matrix C_Y , we can use the following known procedure: Calculate the Cholesky decomposition $C_Y = A A^T$. Generate a vector \vec{x} of uncorrelated random numbers with all $\text{var}(X_i) = 1$. Apply the matrix A: $\vec{Y} = A\vec{X}$. This will result in correlated random variables with covariance matrix $C_Y = A A^T$.
- With the formulas outlined in this Wikibook, we can additionally study the shape of the resulting distribution and the effect of non-linear transformations. Consider, e.g., that X is uniform distributed in $[0, 2\pi]$, $Y_1 = \sin(X)$ and $Y_2 = \cos(X)$. In this case, a 2D plot of random numbers from (Y_1, Y_2) will show a uniform distribution on a circle. Although Y_1 and Y_2 are stochastically dependent, they are uncorrelated. It is therefore important to know the resulting distribution, because $\varrho_{\vec{Y}}(y_1, y_2)$ has more information than the covariance matrix C_Y .

Latest version of this article online on: [https://en.wikibooks.org/wiki/Probability/Transformation of Probability Densities](https://en.wikibooks.org/wiki/Probability/Transformation_of_Probability_Densities)

Author: Lars Winterfeld

Licence: [Creative Commons Attribution-Share Alike 3.0](https://creativecommons.org/licenses/by-sa/3.0/)