Chapter 6

Distribution of a Function of a Random Variable

We want now to discuss how to compute the *p.d.f.* of a function of a random variable. We will start with the simple case of a discrete variable and then we will discuss the case of continuous *p.d.f.*.

6.1 Discrete Variable

We will assume first that there is a one to one correspondence between the values of a variable x and the function $h(x)$. In other words the function $h(x)$ is a single valued. From the definition of probability it follows that the probability for x to have a particular value x_0 is the same as the probability for $h(x)$ to have the value $h(x_0)$, i.e.:

$$P\left(h(x) = h_0\right) = P\left(h(x) = h(x_0)\right) = P(x = x_0) \tag{6.1}$$

Example 6.1 *$h(x) = cos(x)$ (see Tab.6.1) This case is simple since there is a one to one correspondence*

x	0	$\pi/4$	$\pi/2$
$P(x)$	1/6	2/6	3/6
$h(x)$	1	0.71	0
$P(h)$	1/6	2/6	3/6

Table 6.1: *How to compute the probability of a function of a discrete random variable.*

between the values of x and of h and we can use directly equation 6.1.

In the case that there are several x_i which give the same value of h, we have to sum the probabilities:

$$P(h(x) = h_0) = \sum_{i:h(x_i)=h_0} P(x = x_i) \tag{6.2}$$

Formally this can be done since the events $(X = x_i)$ are disjoint:

$$(X = x_i) \bigcap (X = x_j) = \emptyset \; if \; i \neq j$$

and since the event $(H = h_i) = \cup_{some\ k}(X = x_k)$; hence $P(H = h_i) = \sum_{some\ k} P(X = x_k)$

Example 6.2 *$h(x) = cos(x)$, see Tab. 6.2. In this case there is not a one to one correspondence between x and h and we have to fill the table using equation 6.2.*

x	$-\pi/2$	$-\pi/4$	0	$\pi/4$	$\pi/2$
$P(x)$	1/35	3/35	6/35	10/35	15/35
$h(x)$	0	0.71	1.	0.71	0
$P(h)$	16/35	13/35	6/35		

Table 6.2: *The case of discrete random variable for a multi valued function*

6.2 The Case of one continuous Variable

By definition $f(x)dx = P(x \leq X \leq x + dx)$, i.e. $f(x)dx$ is the probability to find the value of X in the interval $(x, x + dx)$. In the case of the variable $h = h(x)$ the probability density is defined as:

$$g(h)dh = P(h \leq H \leq h + dh) = f(x)dx$$

To compute $g(h)$, let us assume first that $h(x)$ is a one-to-one function of x. We will find a small range of values of h which are produced by a small range of values of x with probability $f(x)dx$.

Let

$$x = x(h)$$

be the inverse of the function h. To compute the differential dx as a function of h, we have to take care of its sign, since the *p.d.f.* are NON negative by definition.

$$dx = \left|\frac{\partial x(h)}{\partial h}\right| \, dh$$

The sign of the derivative keep track of the sign of how the variables change. A negative sign indicates that, at x, an increase of x corresponds a decrease of h. Since we are concerned of the probability content of the intervals, the sign is not relevant.

Combining the two equations we get:

$$g(h) = f(x(h)) \cdot \left|\frac{\partial x(h)}{\partial h}\right|$$

Example 6.3 *$h(x) = cos(x)$. Assume that the p.d.f. of x is*

$$f(x) = a + bx \qquad (0 \leq x \leq \pi/2)$$

with $a = 1/\pi$ and $b = 4/\pi^2$.

$$x = ArcCos(h) \qquad \frac{dx(h)}{dh} = -\frac{1}{\sqrt{1-h^2}}$$

We then get ($0 \leq h \leq 1$):

$$g(h)\ dh = f(x(h)) \cdot \frac{\partial x}{\partial h} = \frac{a + b \cdot ArcCos(h)}{\sqrt{1-h^2}}\ dh$$

In the case there are several values of x which give the same value $h(x)$, we have to sum the probabilities, as we did in the discrete case (see Fig.6.1):

$$g(h)\ dh = \sum_{h(x_i)=h} f(x_i(h)) \cdot \left|\frac{\partial x_i}{\partial h}\right| \, dh$$

Example 6.4 *Assume that the variable X has a Standard Normal p.d.f:*

$$X \sim f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

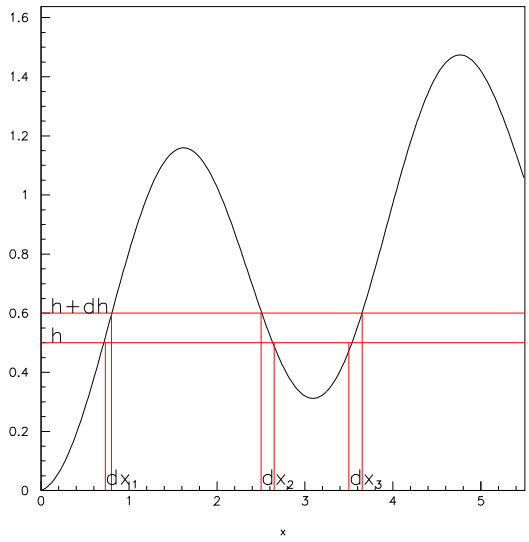

Figure 6.1: *Multi-valued function. To the same interval (h,h+dh) corresponds the intervals* (x_1,x_1+dx_1), (x_2,x_2+dx_2)....

let us compute the p.d.f. of the function $h=x^2$. *The function has two inverses:* $x_1=\sqrt{h}$ *and* $x_2=-\sqrt{h}$. *The absolute values of the derivatives are:*

$$\left|\frac{\partial x_{1,2}}{\partial h}\right| = \frac{1}{2\sqrt{h}}$$

hence, using the previous rules, we get:

$$\begin{aligned} g(h)dh &= \left(f\big(x_1(h)\big)\left|\frac{\partial x_1}{\partial h}\right| + f\big(x_2(h)\big)\left|\frac{\partial x_2}{\partial h}\right| \right) dh \\ &= \frac{1}{\sqrt{2\pi}}\left(e^{-h/2}\frac{1}{2\sqrt{h}} + e^{-h/2}\frac{1}{2\sqrt{h}} \right) dh \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{h}} e^{-\frac{h}{2}}\, dh \end{aligned}$$

The function can be cast in the form:

$$g(h) = \frac{h^{\alpha-1}e^{-h/\beta}}{\Gamma(\alpha)\beta^{\alpha}} \qquad (\alpha=1/2, \quad \beta=2, \quad \Gamma(\frac{1}{2})=\sqrt{\pi}) \tag{6.3}$$

The density is of the family of the Γ.

Example 6.5 *If a continuous random variable* X *has a density* $f(X)$, *we define its distribution as:*

$$F(X) = \int_{-\infty}^{X} f(t)\ dt$$

Let us consider the transformation:

$$X \to Y = F(X) \qquad (0 \le Y \le 1)$$

Let us compute the density of Y. *$F(X)$ is strictly increasing hence:*

$$g(Y)dY = f\left(X(Y)\right)\left|\frac{dX}{dY}\right|\ dY$$

Since $dY/dX = f(X)$ we obtain $g(Y) = 1$. This transformation generates an uniform density, for any density $f(X)$.

Another equivalent proof is the following:

$$G(Y') = P(Y \leq Y') = P(X \leq X') = F(X') = Y'$$

where the second equality follows from the strict monotonicity of distributions and $F(X') = Y'$ by definition. The density of Y is then:

$$g(y) = \frac{dG(Y)}{dY} = 1$$

A more, mathematically, sound derivation of the rules outlined above is the following: assume:

$$X \sim f_X(x) \qquad \text{and} \qquad H = g(x) \qquad \text{a function of } X$$

Let us make use of the definition of distribution:

$$P(H \leq h) = \int_{X:g(x)\leq h} f_X(x)dx = F_H(h)$$

The integral is performed in the region such that $g(x) \leq h$. From the above distribution we get the needed density by:

$$f_H(h) = \frac{d}{dh}F_H(h)$$

Example 6.6 *Let us consider the linear transformation:*

$$\begin{aligned} H &= a + bX = g(x) \qquad \left(\frac{dx}{dh} = \frac{1}{b}\right) \\ F_H(h) &= \int_{a+bx\leq h} f_X(x)dx \end{aligned}$$

The integral is then performed in the two regions:

$$\begin{aligned} x &\leq \frac{h-a}{b} \qquad \text{if } b > 0 \\ x &\geq \frac{h-a}{b} \qquad \text{if } b < 0 \end{aligned}$$

Hence the density is:

$$\begin{aligned} f_H(h) &= \frac{d}{dh}\int_{-\infty}^{(h-a)/b} f_X(x)dx = \frac{1}{b}f_X((h-a)/b) \qquad \text{if } b > 0 \\ f_H(h) &= \frac{d}{dh}\left(1 - \int_{-\infty}^{(h-a)/b} f_X(x)dx = -\frac{1}{b}f_X((h-a)/b)\right) \qquad \text{if } b < 0 \end{aligned}$$

Or, more simply:

$$f_H(h) = \left|\frac{1}{b}\right| f_X((h-a)/b) \qquad \text{if } b \neq 0$$

as we could have deduced from our previous method.

Example 6.7 *Assume $X \sim f_X(x)$ and let us consider the transformation $W = X^2$. The calculation of the density of W proceeds through the calculation of the distribution as we have done above:*

$$F_W(w) = P(W \leq w) = P(X^2 \leq w) = P(|X| \leq \sqrt{w}) = P(-\sqrt{w} \leq X \leq \sqrt{w})$$

We have now make explicite the integrals:

$$F_W(w) = \int_{-\infty}^{+\sqrt{w}} f_X(x)dx - \int_{-\infty}^{-\sqrt{w}} f_X(x)dx$$

Hence the density is:

$$f_W(w) = \frac{1}{2\sqrt{w}} f_X(\sqrt{w}) + \frac{1}{2\sqrt{w}} f_X(-\sqrt{w})$$

If X would be distributed Normally, then the previous formula produces the χ^2_1 distribution, as derived in a previous example.

6.3 The Case of several Continuous Variables

The previous relations can be extended to the case of several variables. Suppose we have:

$$\{x_1, x_2, \cdots, x_n\} \sim f(x_1, x_2, \cdots, x_n)$$

that we want to transform to:

$$\{h_1, h_2, \cdots, h_n\} \qquad\qquad h_i = h_i(x_1, x_2, \cdots, x_n)$$

We have to find the inverse transformations:

$$x_i = x_i(h_1, h_2, \cdots, h_n)$$

The procedure, as in the case of a single variable consists in transforming the variables ($\vec{x} \to \vec{h}$) and in summing over the volumes of $\vec{x}$ that transform to the same volume in $\vec{h}$:

$$g(h_1 \cdots h_n) = \sum_{\vec{x_i}:\left(\vec{h}(\vec{x_i})=h\right)} \left(f(x_1(\vec{h}) \cdots x_n(\vec{h})) \cdot \left| \frac{\partial(x_1 \cdots x_n)}{\partial(h_1 \cdots h_n)} \right| \right)$$

where the last term is the absolute value of the Jacobian:

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial h_1} & \frac{\partial x_1}{\partial h_2} & \cdots & \frac{\partial x_1}{\partial h_n} \\ \frac{\partial x_2}{\partial h_1} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial h_1} & \cdots & \cdots & \frac{\partial x_n}{\partial h_n} \end{bmatrix}$$

Example 6.8 *Let us consider the simple case of only two variables, A and B with p.d.f. $P_A(A)$ and $P_B(B)$ and let us assume that they are not correlated. The joint p.d.f. is:*

$$P_{AB}(A, B) = P_A(A) \cdot P_B(B)$$

We have to compute the p.d.f. of a function of A and B: $C = f(A; B)$. We have to consider two new variables:

$$\begin{cases} C &= f(a, b) \\ G &= B \end{cases}$$

We have now to invert the previous relation in order to write:

$$\begin{cases} A &= A(c, g) \\ B &= G \end{cases}$$

The Jacobian is:

$$J = \frac{\partial(A, B)}{\partial(C, G)} = \begin{vmatrix} \frac{\partial A}{\partial c} & \frac{\partial A}{\partial g} \\ \frac{\partial B}{\partial c} & \frac{\partial B}{\partial g} \end{vmatrix} = \frac{\partial A}{\partial c}$$

The joint density of c and g is:

$$H(c, g) = P_A(a(c, g)) \cdot P_B(g) \, |J|$$

The density of c is obtained integrating away the variable g:

$$C(c) = \int_{-\infty}^{\infty} H(c, g) \, dg$$

All this can be rather complicate.

Example 6.9 *Let us make an example of the procedure outlined in the previous example. Assume that*

$$\begin{cases} a & \sim & N(0,1) \\ b & \sim & N(\mu,1) \end{cases}$$

And we are interested in the function $c = a + b$.

$$\begin{cases} c & = & a+b \\ g & = & b \end{cases} \qquad \begin{cases} a & = & c-g \\ b & = & g \end{cases}$$

The Jacobian is one and the joint density is:

$$H(c,g) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}} e^{-(c-g)^2/2} e^{-(\mu-g)^2/2}$$

Notice that the joint density cannot be factorized in two terms functions only of c and g. This means that the variables c and g are not independent.

The integral is performed squaring the exponent:

$$H(c,g) = \frac{1}{2\pi} exp\left[-\frac{1}{2}\left(\frac{(c-\mu)^2}{2} + \frac{(2g-(c+\mu))^2}{2}\right)\right]$$

and, integrating away g we obtain:

$$\begin{aligned} C(c) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} exp\left[-(g-\frac{c+\mu}{2})^2\right]\, dg\, exp\left[-\frac{(c-\mu)^2}{4}\right] \\ &= \frac{1}{\sqrt{2}\sqrt{2\pi}} e^{-\frac{(c-\mu)^2}{4}} \end{aligned}$$

which is, as we could have anticipated, a Normal distribution with mean μ and variance $1+1=2$. The procedure could become rather complicate. In many cases the complete calculation of the probability density is not needed; often only the mean and the variance are important and linearization or numerical calculation can solve the problem.

Example 6.10 *The two independent random variables x and y are normal standard:*

$$\begin{cases} x & \sim & N(0,1) \\ y & \sim & N(0,1) \end{cases}$$

we are asked to compute the density of $u = x/y$. We proceed as in the previous examples by changing the variable:

$$\begin{cases} u & = \frac{x}{y} \\ v & = y \end{cases} \qquad \begin{cases} x & = u \cdot v \\ y & = v \end{cases}$$

The Jacobian is:

$$|J| = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \begin{vmatrix} v & 0 \\ u & 1 \end{vmatrix} = |v|$$

Hence:

$$h(u,v)\, dudv = f(x)\cdot f(y)\, dxdy = f(uv)\cdot f(v)\,|J|\, dudv = \frac{1}{2\pi} e^{-(uv)^2/2} e^{-v^2/2}\,|v|\, dudv$$

To obtain the p.d.f. we have now to integrate on v:

$$U(u) = \frac{1}{2\pi}\left(\int_{-\infty}^{0} (-v) e^{-v^2(1+u^2)/2} dv + \int_{0}^{+\infty} (v) e^{-v^2(1+u^2)/2} dv\right) = \frac{1}{\pi}\frac{1}{1+u^2}$$

and we have obtained the Cauchy distribution.

6.4 Linearized Approximation and the "Propagation of errors"

This section describes a method very often used in the practice when we want to compute the standard deviation of a function of a random variable.

In principle we should know or have derived the *p.d.f.* of the new variable (z) with the methods of the previous section and then compute:

$$\mu_z = E(z(x)) \qquad \sigma_z^2 = E((z(x) - \mu_z)^2)$$

However the procedure is often long and it is usually adequate to linearize the function and use the previous theorems.

Assume we have measured random variables

$$x_1, x_2, \cdots, x_n = \vec{x}$$

and we want to compute the variance of a function of these variables $y(\vec{x})$. We assume that we can perform a Taylor expansion around μ and truncate to the first term:

$$y(\vec{x}) = y(\vec{\mu}) + \sum_{i=1,n} (x_i - \mu_i) \left.\frac{\partial y}{\partial x_i}\right|_{\vec{x}=\vec{\mu}} + \cdots \tag{6.4}$$

If we take the expectation value, the first order term vanishes:

$$\begin{aligned} E(y(\vec{x})) &= y(\vec{\mu}) + \cdots \\ Var(y(\vec{x})) &= E((y - E(y))^2) = \\ &= E(y(\vec{x}) - y(\vec{\mu}))^2 = \\ &= \sum_{i=1,n}\sum_{j=1,n} \left.\left(\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right)\right|_{\vec{x}=\vec{\mu}} E((x_j - \mu_j)(x_i - \mu_i)) \end{aligned}$$

where, by definition:

$$E((x_j - \mu_j)(x_i - \mu_i)) = Var(\vec{x})|_{i,j} = \begin{bmatrix} Var(x_1) & Cov(x_1, x_2) & \cdots \\ Cov(x_2, x_1) & Var(x_2) & \cdots \\ \cdots & \cdots & Var(x_n) \end{bmatrix}_{i,j}$$

is the covariance matrix of $\vec{x}$.

In the approximation discussed above the variance of y is, in general, given by:

$$\begin{aligned} Var(y) &= \sum_{i=1,n}\sum_{j=1,n} \left.\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right|_{\vec{x}=\vec{\mu}} \cdot (Var(\vec{x}))_{i,j} \\ &= \sum_{i=1}^{n} \sigma_i^2 \left.\left(\frac{\partial y}{\partial x_i}\right)^2_{\vec{x}=\vec{\mu}} + 2\sum_{i>j} Cov(x_i, x_j) \left.\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right|_{\vec{x}=\vec{\mu}} \end{aligned}$$

If the measured variables x_i are uncorrelated, the variance matrix is diagonal and the previous equation simplifies to:

$$Var(y) = \sum_{i=1,n} \left.\left(\frac{\partial y}{\partial x_i}\right)^2\right|_{\vec{x}=\vec{\mu}} \cdot \sigma_i^2(\vec{x})$$

This is the well known formula for propagating the errors. In fact in most of the cases the x_i are uncorrelated (or their correlation is neglected). The general formula has to be used in the case the variables are correlated.

Example 6.11 *Let us consider the function $Z = X \cdot Y$. The random variables X and Y have unspecified densities with mean and variances: μ_X,μ_Y,σ_X^2,σ_Y^2,$\sigma_{X,Y}^2$. An approximate evaluation of the variance of*

Z is:

$$
\begin{aligned}
Var(Z) \approx & \left(\frac{\partial Z}{\partial X}\right)^2_{\mu_Y} \cdot \sigma_X^2 + \left(\frac{\partial Z}{\partial Y}\right)^2_{\mu_X} \cdot \sigma_Y^2 + 2 \left.\frac{\partial Z}{\partial X}\frac{\partial Z}{\partial Y}\right|_{\mu_X,\mu_Y} \cdot \sigma_{X,Y}^2 \\
= & \mu_Y^2\sigma_X^2 + \mu_X^2\sigma_Y^2 + 2\ \mu_X\mu_Y\sigma_{X,Y}^2 \\
\frac{Var(Z)}{\mu_X^2\mu_Y^2} \approx & \left(\frac{\sigma_X}{\mu_X}\right)^2 + \left(\frac{\sigma_Y}{\mu_Y}\right)^2 + 2\ \frac{\sigma_{X,Y}}{\mu_X\mu_Y}
\end{aligned}
$$

Now one should ask: when is the linearization allowed? There is no general answer. Figure 6.2 shows an example of what happens. The variable x is distributed as a Normal ($N(\mu,\sigma^2)$) and the variable $y = x^2$. A measure of the variability of x is the standard deviation σ. The function y can be linearized locally around μ (the line marked as *Lin. Appr.*). If, in the region of variability of x, the function y is well represented by the straight line, the linearization is good. This depends upon the value of σ as well as on the value of local curvature of the function.

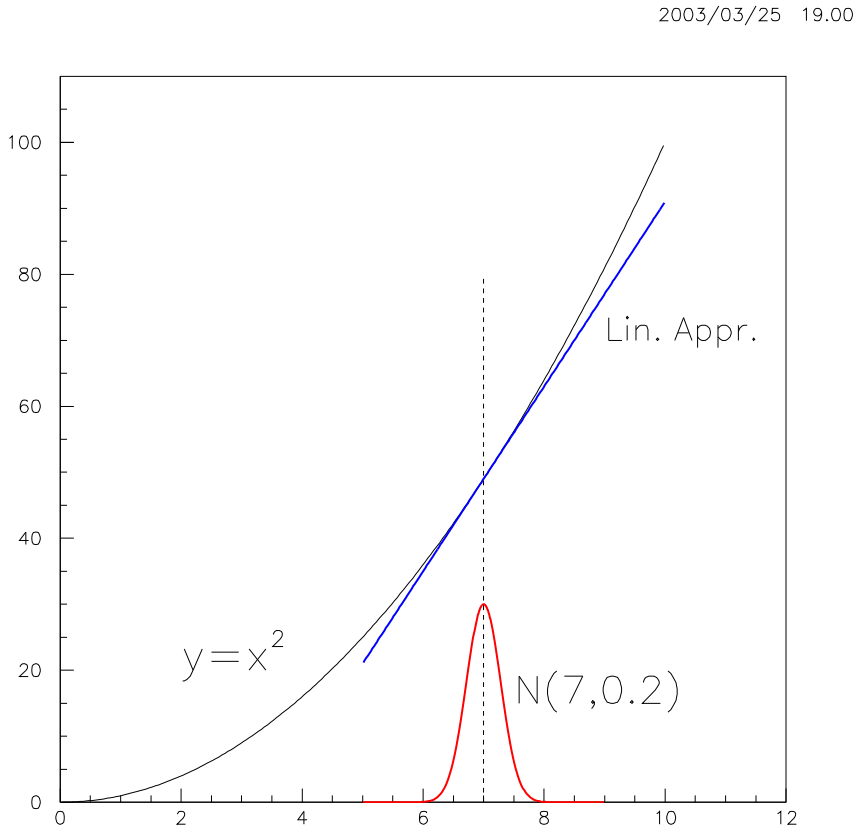

Figure 6.2: *An example of how the linearization works. The random variable x $N(\mu,\sigma^2)$ and the function of x is $y = x^2$.*

Example 6.12 *Let us consider the function $Z = X \cdot Y$. The random variables X and Y are independent and have unspecified densities with mean and variances: μ_X,μ_Y,σ_X^2,σ_Y^2. In this case we can compute exactly the variance of the variable Z:*

$$
\begin{aligned}
Var(Z) = & E\left((XY - \mu_X\mu_Y)^2\right) = E\left((X\cdot Y)^2\right) - (\mu_X\cdot\mu_Y)^2 = E(X^2)\cdot E(Y^2) - (\mu_X\cdot\mu_Y)^2 \\
= & \left(\sigma_X^2 + \mu_X^2)\right)\left(\sigma_Y^2 + \mu_Y^2)\right) - (\mu_X\cdot\mu_Y)^2 \qquad \textit{In fact: } \sigma_X^2 = E(X^2) - \mu_X^2 \\
\frac{Var(Z)}{(\mu_X\mu_Y)^2} = & \frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_X^2\sigma_Y^2}{\mu_X^2\mu_Y^2}
\end{aligned}
$$

The linear approximation holds if the last term at the right side can be neglected, i.e. if either $\frac{\sigma_X^2}{\mu_X^2} \ll 1$ *or* $\frac{\sigma_Y^2}{\mu_Y^2} \ll 1$.

Exercise 6.1 *Compute the second order correction to* $E(Y)$ *in equation 6.4and discuss the approximations of the linear method.*

Exercise 6.2 *Assume that* x_1 *and* x_2 *are uncorrelated random variables having variances* $Var(x_1)$ *and* $Var(x_2)$ *respectively. Compute the variance of the functions:* $y = x_1 \cdot x_2$ *and* $y = \frac{x_1}{x_2}$*, assuming that we can use a linear approximation.*

The inverse of the covariance matrix is called the *weight matrix.*

Exercise 6.3 *Assume* $x_1, x_2, \cdots, x_n$ *are* n *measurements of the same quantity. It is convenient to consider the* x_i *as independent variables. Assume that all the measurements have the same variance* $Var(x) = \sigma^2$*. Compute the variance of the function* $y = \frac{\sum x_i}{n}$.

The previous formalism can be generalized to the case where a set of functions $\vec{y}$ is function of $\vec{x}$.

$$y_j(x_i), \qquad\qquad i = 1...n, \;\; j = 1...m$$

Again we shall follow the approximate procedure by linearizing the functions by Taylor expansion about the mean.

$$y_j(\vec{x}) = y_j(\vec{\mu}) + \sum_i (x_i - \mu_i) \frac{\partial y_j}{\partial x_i}\bigg|_{\vec{x}=\vec{\mu}} + \cdots \tag{6.5}$$

The variance of $\vec{y}$ is:

$$\begin{aligned} Var(y)_{k,l} &= E\left((y_k - E(y_k))(y_l - E(y_l))\right) = \\ &= \sum_{i=1,n} \sum_{j=1,n} \left(\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j}\right)\bigg|_{\vec{x}=\vec{\mu}} E((x_i - \mu_i)(x_j - \mu_j)) = \\ &= \sum_{i=1,n} \sum_{j=1,n} \left(\frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j}\right)\bigg|_{\vec{x}=\vec{\mu}} Var_{i,j}(\vec{x}) \end{aligned}$$

It convenient to express these results in matrix notation.

$$\vec{y} = \vec{\mu_y} + S \cdot (\vec{x} - \vec{\mu}) + \cdots$$

S is a matrix of the derivatives:

$$S_{k,l} = \frac{\partial y_k}{\partial x_l} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots \\ \cdots & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}_{m \times n}$$

and the law of propagation of errors becomes:

$$Var(y)_{(m\times m)} \approx S_{(m\times n)} Var(x)_{(n\times n)} S^T_{(n\times m)} \tag{6.6}$$

We have to stress that the formulas are approximate and their validity holds till we can neglect the non linear terms of the functions. In the general case we have to use the full machinery by computing the *p.d.f.* of the new variable or use MC methods.

If y_i are linear functions of x_j then equation 6.5 is exact and also equation 6.6.

Summary of this Chapter

Change of variable

If: $x \sim f(x)$ and $h = h(x)$ then:
This can be generalized:
if $\{x_1, \cdots, x_n\} \sim f(x_1, \cdots, x_n)$
($\vec{x} \sim f(\vec{x})$ in short) and
$\begin{cases} h_j = f_j(\vec{x}) \\ j = 1, \cdots n \end{cases}$

$$h \sim g(h) = \sum_{h(x_i)=h} f\big(x_i(h)\big) \left|\tfrac{\partial x(h)}{\partial h}\right|_{x_i}$$

$$\vec{h} \sim \sum_{h(\vec{x}_i)=\vec{h}} f\left(\vec{x}_i(\vec{h})\right) \left|\tfrac{\partial\{x_1 \cdots x_n\}}{\partial\{h_1 \cdots h_n\}}\right|_{x_i}$$

An alternative procedure is through the distribution:

$$H(w) = P(h \leq w) = \int \cdots \int_{\vec{x} \in A} f\left(\vec{x}(\vec{h})\right) d^n x$$

$$A : \vec{h}(\vec{x}) \leq w$$

$$h(w) = \frac{\partial H(w)}{\partial w}$$

The linear approximation

To compute mean and variance a linear approximation can be used:
$\begin{cases} h_j(\vec{x}) \approx h_j(\vec{\mu}) + \sum (x_i - \mu_i) \frac{\partial h_j}{\partial x_i}\big|_{\vec{x}=\vec{\mu}} + \cdots \\ j = 1, \cdots m \end{cases}$

$$\begin{cases} E(h) & \approx h(E(\vec{x})) = h(\vec{\mu}) \\ Var(h) & \approx G_h \cdot Cov(x) G_h^T \end{cases}$$

where

$$\begin{cases} G_h\big|_{i,j} & = \partial_{x_i} h_j \\ Cov(x)\big|_{i,j} & = E\big((x_i - \mu_i)(x_j - \mu_j)\big) \end{cases}$$

$$\boxed{Var(y) = \sum_{i=1}^{n} \sigma_i^2 \left(\frac{\partial y}{\partial x_i}\right)^2_{\vec{x}=\vec{\mu}} + 2 \sum_{i>j} Cov(x_i, x_j) \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j}\bigg|_{\vec{x}=\vec{\mu}}}$$

If the measurements are uncorrelated:

$$\begin{cases} Cov(x)\big|_{i,j} & = \sigma_i^2 \delta_{i,j} \\ Var(h) & = \sum \big(\partial_{x_i} h\big)^2 Var(x_i) \end{cases}$$

$$\boxed{Var(y) = \sum_{i=1}^{n} \sigma_i^2 \left(\frac{\partial y}{\partial x_i}\right)^2_{\vec{x}=\vec{\mu}}}$$