The Jacobian

Let $\mathbf{h} = (h_1, \ldots, h_k)$ where each h_i is a real-valued function on R^k. Define the Jacobian $J_{\mathbf{h}}(\mathbf{t})$ of **h** evaluated at $\mathbf{t} = (t_1, \ldots, t_k)$ to be

$$J_{\mathbf{h}}(\mathbf{t}) = \begin{vmatrix} \frac{\partial}{\partial t_1} h_1(\mathbf{t}) & \cdots & \frac{\partial}{\partial t_1} h_k(\mathbf{t}) \\ \vdots & & \\ \frac{\partial}{\partial t_k} h_1(\mathbf{t}) & \cdots & \frac{\partial}{\partial t_k} h_k(\mathbf{t}) \end{vmatrix}.$$

We have the following change of variable theorem for multiple integrals from calculus. For a proof of this result see a text on advanced calculus or mathematical analysis.

Theorem: Let $\mathbf{h} = (h_1, \ldots, h_k)$ be a transformation defined on an open subset *B* of R^k. Suppose that

i) **h** has continuous first partial derivatives on *B*.
ii) **h** is one-to-one on *B*.
iii) The Jacobian of **h** is nonzero everywhere on *B*.

Let *f* be a real-valued function on the range $\mathbf{h}(B) = \{(h_1(\mathbf{t}), \ldots, h_k(\mathbf{t})) : \mathbf{t} \in B\}$ of **h** and suppose

$$\int_{\mathbf{h}(B)} |f(\mathbf{x})| d\mathbf{x} < \infty.$$

Then, for every measurable subset *K* of $\mathbf{h}(B)$

$$\int_K f(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{h}^{-1}(K)} f(\mathbf{h}(\mathbf{t})) |J_{\mathbf{h}}(\mathbf{t})| d\mathbf{t}.$$

Note that $d\mathbf{x} = dx_1 \cdots dx_k$ and $\mathbf{h}^{-1}(\mathbf{x}) = \mathbf{t}$ if and only if $\mathbf{x} = \mathbf{h}(\mathbf{t})$. We also know from calculus that

$$J_{\mathbf{h}^{-1}}(\mathbf{t}) = \frac{1}{J_{\mathbf{h}}(\mathbf{h}^{-1}(\mathbf{t}))}.$$

We can now prove the following theorem:

Theorem: Let **X** be a continuous random vector and let S be an open subset of R^k such that $P(\mathbf{X} \in S) = 1$. If $\mathbf{g} = (g_1, \ldots, g_k)$ is a transformation from S to R^k such that **g** and S satisfy the conditions of the above theorem, then the density of $\mathbf{Y} = \mathbf{g}(\mathbf{X})$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(\mathbf{g}^{-1}(\mathbf{y})\right)\left|J_{\mathbf{g}^{-1}}(\mathbf{y})\right|$$

for $\mathbf{y} \in \mathbf{g}(S)$.

Proof: The distribution of **Y** is given by

$$F_{\mathbf{Y}}(\mathbf{y}) = \int_{-\infty}^{y_k} \cdots \int_{-\infty}^{y_1} f_{\mathbf{Y}}(y_1, \ldots y_k) dy_1 \cdots dy_k = \int \underset{A_k}{\cdots} \int f_{\mathbf{X}}(x_1, \ldots x_k) dx_1 \cdots dx_k$$

where, $A_k = \left\{\mathbf{x} \in R^k : g_i(\mathbf{x}) \le y_i,\ i = 1, \ldots, k\right\}$. We now apply the above theorem with $\mathbf{h} = \mathbf{g}^{-1}$ and $f = f_{\mathbf{X}}$. Since

$$\mathbf{h}^{-1}(A_k) = \mathbf{g}(A_k) = \left\{\mathbf{g}(x) : g_i(\mathbf{x}) \le y_i, i = 1, \ldots, k\right\} = \left\{\mathbf{t} : t_i \le y_i,\ i = 1, \ldots, k\right\}$$

we get

$$F_{\mathbf{Y}}(\mathbf{y}) = \int_{-\infty}^{y_k} \cdots \int_{-\infty}^{y_1} f_{\mathbf{X}}\left(\mathbf{g}^{-1}(\mathbf{t})\right)\left|J_{\mathbf{g}^{-1}}(\mathbf{t})\right| dt_1 \cdots dt_k \Rightarrow f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(\mathbf{g}^{-1}(\mathbf{y})\right)\left|J_{\mathbf{g}^{-1}}(\mathbf{y})\right|.$$

Now, this last theorem required **g** to be one-to-one. We can generalize this when **g** is not one-to-one as follows. Let $S_1, \ldots, S_r$ be r disjoint subsets of R^n such that $P\left(\mathbf{X} \in \bigcup_{i=1}^r S_i\right) = 1$. Suppose that **g** is a transformation from $\bigcup_{i=1}^r S_i$ to R^n such that

i) **g** has continuous first partial derivatives on S_i for each i.
ii) **g** is one-to-one on each S_i.
iii) The Jacobian of **g** is nonzero everywhere on each S_i.

Then,

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^{r} f_{\mathbf{X}}\left(\mathbf{g}_i^{-1}(\mathbf{y})\right)\left|J_{\mathbf{g}_i}\left(\mathbf{g}_i^{-1}(\mathbf{y})\right)\right|^{-1} I_i(\mathbf{y}),\ \mathbf{y} \in g\left(\bigcup_{i=1}^r S_i\right)$$

where, $\mathbf{g}_i$ is the restriction of **g** to S_i and $I_i(\mathbf{y}) = 1$ if $\mathbf{y} \in \mathbf{g}(S_i)$ and equals 0 otherwise. If $I_i(\mathbf{y}) = 0$ then the whole summand is taken to be 0 even though $\mathbf{g}_i^{-1}$ is undefined there.

This more general theorem is proved using $P\left(\mathbf{g}(\mathbf{X}) \in B\right) = \sum_{i=1}^{r} P\left(\mathbf{g}(\mathbf{X}) \in B, \mathbf{X} \in S_i\right)$. The proof of this more general result now follows along the same lines as the proof of the last theorem so the details will be omitted.

The results presented above are taken from Bickel and Doksum, *Mathematical Statistics*, Prentice Hall, 1977.