14. CHANGE OF VARIABLE - I

Let X be a random variable on a probability space. Let $T:\mathbb{R}\to\mathbb{R}$ be a given function. Let $Y := T(X)$. Then Y is a random variable on the same probability space as X.

Question: If we know the distribution of X, how to calculate the distribution of Y?

Answer: First, let us take the case of a discrete random variable X. Assume that $\mathcal{R}(X)=\{x_1,x_2,\ldots\}$ and $\mathbf{P}\{X=x_j\}=p_j$. Then, the range of Y is $\{T(x_1),T(x_2),\ldots\}$. Of course these are not al distinct, for example we could have $T(x_1)=T(x_2)$. Hence, let us write the set as $\{y_1,y_2,\ldots\}$, omitting repetition. Then, $\mathbf{P}\{Y=y_j\}=\mathbf{P}\{T(X)=y_j\}=\sum_{i:T(x_i)=y_j} p_i$. Thus, from the pmf of X, we can calculate the pmf of Y.

The same idea can be pursued to give a solution to the question for general X and T. For any set $A\subseteq\mathbb{R}$, $\mathbf{P}\{Y\in A\}=\mathbf{P}\{X\in B\}$, where $B=\{u\in\mathbb{R}: T(u)\in A\}$. If we knew the distribution of X, we could calculate the $\mathbf{P}\{X\in B\}$ in principle. Thus, we could calculate $\mathbf{P}\{Y\in A\}$ for any A, and thus we would know everything about the distribution of Y.

However, finding the set B given the set A could be tedious and in general difficult. Thus, the above solution is often completely useless in practice. Instead, we shall restrict ourselves to two special situations and find answers that are actually useful in practice.

(1) Assume that T is strictly increasing and continuous. We make no assumptions on X.

Then, as T is increasing, $B := \lim_{u\to+\infty} T(u)$ and $A := \lim_{u\to-\infty} T(u)$ exist and $A<B$. Further, as T is continuous, for any $x\in(A,B)$, there is some $u\in\mathbb{R}$ such that $T(u)=x$ (this is obvious if you draw the graph of T. A rigorous proof of such things are relegated to the Analysis-I class). Since T is also strictly increasing, T is one-one. The upshot is that $T:\mathbb{R}\to(A,B)$ is one-one and onto and hence must have an inverse $T^{-1}:(A,B)\to\mathbb{R}$ (also one-one and onto).

Now let $t\in(A,B)$ and consider

$$\mathbf{P}\{Y\le t\}=\mathbf{P}\{T(X)\le t\}=\mathbf{P}\{X\le T^{-1}(t)\}=F_X(T^{-1}t)$$

where F_X is the CDF of X. Thus, the CDF of Y is given by $F_Y(t)=F_X(T^{-1}t)$ for $t\in(A,B)$. Of course, for $t\le A$, $F_Y(t)=0$ and for $t\ge B$ we have $F_Y(t)=1$.

Example 47. Let $X\sim \text{Exp}(1)$ and let $T(u)=2u$ so that $Y=2X$. Then $F_X(u)=1-e^{-u}$ (for $u>0$) and hence $F_Y(t)=F_X(T^{-1}t)=F_X(t/2)=1-e^{-t/2}$ which shows that $Y\sim\text{Exp}(1/2)$. More generally one can check in the same way that if $X\sim\text{Exp}(\lambda)$ then $\theta X\sim\text{Exp}(\lambda/\theta)$ for any $\theta>0$ and any $\lambda>0$.

(2) Assume that T is strictly increasing and continuous and also has a (piecewise) continuous derivative. We also assume that X has a density f. With $Y=T(X)$, we want to ask if Y has a pdf and if so how to find it.

Before giving the answer, let us say why we need strict increase of Y. If $T(u)=0$ for all u, then $Y=0$ with probability 1 and hence Y does not have density. For Y to have density, we need to assume strict increase. Similarly, the existence of a derivative is essential although we cannot now give an example to say why.

Now return to the assumption. As seen earlier, $F_Y(t)=F_X(T^{-1}t)$ for $t\in(A,B)$ (A,B are as before). In particular, if we differentiate we get

$$F_Y'(t)=F_X'(T^{-1}t)(T^{-1})'(t)=f(T^{-1}t)(T^{-1})'(t).$$

Thus the pdf of Y is given by $g(t)=f(T^{-1}t)(T^{-1})'(t)$.

(3) If T were assumed to strictly decreasing and having a (piecewise) continuous derivative, then the same steps would show that $g(t)=-f(T^{-1}t)(T^{-1})'(t)$ (note that T^{-1} is also decreasing and hence $(T^{-1})'(t)$ is negative and hence $g\ge 0$ as required.

The two cases of increasing and decreasing can be combined to write

$$g(t)=f(T^{-1}t)|(T^{-1})'(t)|.$$

This is often referred to as the *change of variable formula*.

Example 48. With $X\sim\text{Exp}(1)$ and $T(u)=2u$ (so $Y=2X$), we get

$$g(t)=f(t/2)(1/2)=\frac{1}{2}e^{-t/2}$$

showing again that Y has $\text{Exp}(1/2)$ distribution.

Example 49. Let $X \sim N(0,1)$ and let $T(u) = au + b$ where $a > 0$ and $b \in \mathbb{R}$. Then, $T^{-1}(t) = (t-b)/a$ and $(T^{-1})'(t) = 1/a$. We know that $f(u) = e^{-u^2/2}/\sqrt{2\pi}$. Thus $Y = aX + b$ has density

$$g(t) = f((t-b)/a)\frac{1}{a} = \frac{1}{a\sqrt{2\pi}}e^{-\frac{(t-b)^2}{2a^2}}.$$

Thus $Y \sim N(b, a^2)$. More generally, if $X \sim N(\mu, \sigma^2)$, check that $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Example 50. Let $X \sim N(0,1)$ and let $T(u) = u^2$. In this case $Y = X^2$. A problem in applying the change of variable formula is that T is not one-one. Any $t > 0$ has two possible inverse images $\pm\sqrt{t}$. How to apply the change of variable formula here?

Example 51. Let $X \sim \text{Cauchy}_1$ and let $T(u) = 1/u$ for $u \neq 0$. In this case, T is undefined at zero, is decreasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$, but not decreasing on the whole line. Does the change of variable formula for densities apply here?

We give an extension of the change of variable and apply it to the above two examples.

Extended version of change of variable formula: Let X be a random variable with pdf f. Let $T : \mathbb{R} \to \mathbb{R}$ be a function (in factit is sufficient if T is defined on $\mathcal{R}(X)$ only). Assume that there are points $a_1 < a_2 < \ldots < a_m$ such that T is one-one and strictly monotone (decreasing or increasing) on each of the intervals $I_1 = (-\infty, a_1)$, $I_2 = (a_1, a_2), \ldots (a_m, +\infty)$ and has a (piecewise) continuous derivative on each of these intervals.

Let T_k be the restriction of T to I_k. In other words, define functions $T_1 : I_1 \to \mathbb{R}$, $T_2 : I_2 \to \mathbb{R} \ldots T_m : I_m \to \mathbb{R}$ as $T_i(u) = T(u)$. Then each T_i has a range (A_i, B_i) and an inverse function $T_i^{-1} : (A_i, B_i) \to I_i$. The extended change of variable formula says that Y has a density

$$g(t) = \sum_{i=1}^{m} f(T_i^{-1}t)|(T_i^{-1})'(t)|.$$

Here it is understood that if $t \notin (A_i, B_i)$ for some i, then $T_i^{-1}(t)$ is not well-defined and the corresponding summand is dropped from the right hand side.

With this extended principle, let us revisit the two examples again.

Example 52. Let $X \sim N(0,1)$ and let $T(u) = u^2$. In this case $Y = X^2$. Let $I_1 = (-\infty, 0)$ and $I_2 = (0, \infty)$ (in the above notation, $m = 1$ and $a_1 = 0$). Then T is one-one on I_1 and on I_2. Further, $T_1(u) = -\sqrt{u}$ and $T_2(u) = \sqrt{u}$ and both T_1 and T_2 have the range $(0, \infty)$. Clearly $(T_1^{-1})'(t) = -\frac{1}{2\sqrt{t}}$ and $(T_2^{-1})'(t) = \frac{1}{2\sqrt{t}}$ for all $t > 0$. Thus, for any $t > 0$, we have

$$g(t) = f(T_1^{-1}t)|(T_1^{-1})'(t)| + f(T_2^{-1}t)|(T_2^{-1})'(t)| = f(\sqrt{t})\frac{1}{2\sqrt{t}} + f(-\sqrt{t})\frac{1}{2\sqrt{t}} = \frac{1}{\sqrt{2\pi}}e^{-t/2}t^{-1/2}.$$

This shows that $X^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$.

Example 53. Let $X \sim \text{Cauchy}_1$ and let $T(u) = 1/u$ for $u \neq 0$. In this case, again we take $I_1 = (-\infty, 0)$ and $I_2 = (0, \infty)$. T_1 has range $(-\infty, 0)$ while T_2 has range $(0, \infty)$. For $t > 0$, $T_2^{-1}(t) = \frac{1}{t}$ and for $t < 0$ we have $T_1^{-1}(t) = \frac{1}{t}$. Thus we get for $t > 0$

$$g(t) = f(T_2^{-1}(t))|(T_2^{-1})'(t)| = \frac{1}{\pi(1 + \frac{1}{t^2})}\frac{1}{t^2} = \frac{1}{\pi(t^2 + 1)}.$$

Similarly, for $t < 0$ also we get the same answer. Thus, $1/X$ again has Cauchy distribution.

Here are some exercises to try.

Exercise 54. If $U \sim \text{Unif}([0,1])$ find the distribution of

(1) $Y = aU + b$ where $a > 0$ and $b \in \mathbb{R}$.
(2) $1 - U$.
(3) $-\log U$.
(4) U^m where $m \geq 1$ is an integer.

Exercise 55. If $V \sim \text{unif}([-\pi/2, \pi/2])$, then find the distribution of $\sin(V)$.

Exercise 56. If $X \sim \text{Bin}(n, p)$, find the distribution of $n - X$. If $X \sim \text{Hypergeo}(b, w, m)$, find the distribution of $m - X$.

Exercise 57. Let X have pdf f. Find the densities of $X + b$ for $b \in \mathbb{R}$ and aX where $a > 0$ in terms of f (you may also combine these two and just find the density of $aX + b$).

16. A BRIEF REVIEW OF MULTI-VARIABLE CALCULUS

Let us recall a few basic notions from multivariable calculus. Consider a function $T:\mathbb{R}^n\to\mathbb{R}^n$ mapping $\mathbf{x}$ to $(T_1(\mathbf{x}),\dots,T_n(\mathbf{x}))$ where $T_i:\mathbb{R}^n\to\mathbb{R}$. We say that T is continuously differentiable if the partial derivatives $\partial_i T_j(\mathbf{x})$ exist and are continuous. In this case, the derivative of T is the $n\times n$ matrix given by $DT(\mathbf{x})=(\partial_i T_j(\mathbf{x}))$. Every differentiable function can be approximated to the first order by a linear function

$$T(\mathbf{x}+\mathbf{h})=T(\mathbf{x})+DT(\mathbf{x})\mathbf{h}+o(\|\mathbf{h}\|),\quad \text{as } \mathbf{h}\to\mathbf{0}.$$

This means that in a neighbourhood of $\mathbf{x}$, to a first approximation T is given by the linear map $L(\mathbf{h})=\mathbf{b}+A\mathbf{h}$ where $\mathbf{b}=T(\mathbf{x})$ and $A=DT(\mathbf{x})$.

Example 64. Let $T(\mathbf{x})=B\mathbf{x}$ where B is an $n\times n$ matrix. Then, $DT(\mathbf{x})=B$ for all $\mathbf{x}$. The unit cube $[0,1]^n$ transforms under T to the parallelopiped formed by $B\mathbf{e}_1,\dots,B\mathbf{e}_n$ (which are precisely the columns of B). The volume of the latter body is exactly $|\det(B)|$. Thus, in this case, T enlarges the volume of any set by the factor $|\det(B)|$. If this example is not fully clear, start with a diagonal matrix B and then consider general 2×2 matrices.

The usual rules like $D(T+S)=DT+DS$ etc hold as does the *chain rule* $D(S\circ T)(\mathbf{x})=DS(T\mathbf{x})DT(\mathbf{x})$.

Volume change: How do volumes change under a continuously differentiable transformation? This will be of great importance to us. We start from simple cases and go on to deeper ones.

(1) Let $T(\mathbf{x})=B\mathbf{x}$ where $B=\mathrm{diag}(b_1,\dots,b_n)$ is a diagonal matrix. Assume that $b_i>0$ for simplicity. How does T transform volumes? In other vords, if $K\subseteq\mathbb{R}^n$ is a subset, what is $\mathrm{vol}(T(K))$? For example, if $K=[0,1]^n$, then $T(K)=[0,b_1]\times[0,b_2]\times\ldots\times[0,b_n]$ showing that $\mathrm{vol}(T(K))=\mathrm{vol}(K)\det(B)$. It is not difficult to see that this holds for any K. Further, if we allow $b_i\le 0$, then we simply get $\mathrm{vol}(T(K))=\mathrm{vol}(K)|\det(B)|$.

(2) Let $T(\mathbf{x})=B\mathbf{x}$ where B is any $n\times n$ matrix with real entries. Then $T(K)$ is not necessarily a rectangle but a parallelopiped spanned by the vectors $B\mathbf{e}_1,B\mathbf{e}_2,\dots,B\mathbf{e}_n$ where $\mathbf{e}_i$ are the standard co-ordinate vectors. Observe that $B\mathbf{e}_i$ is just the i^{th} column of the matrix B. The volume of the parallelopiped formed by the columns of B is precisely $|\det(B)|$. Thus again we get $\mathrm{vol}(T(K))=\mathrm{vol}(K)|\det(B)|$ (and again it is not hard to see that this generalizes for any K).

(3) Now let T be any continuously differentiable and one-one map. Fix $\mathbf{x}\in\mathbb{R}^n$ and consider the linear map $L(\mathbf{h})=DT(\mathbf{x})\mathbf{h}$ so that $T(\mathbf{x}+\mathbf{h})=T(\mathbf{x})+L(\mathbf{h})+o(\|\mathbf{h}\|)$ for small $\mathbf{h}$. The map L increases volumes by a factor of $\det(DT(\mathbf{x}))$. Thus, if $\delta>0$ is small and $K_\delta=B(\mathbf{x},\delta)$, then $\mathrm{vol}(T(K_\delta))$ is "nearly" $|\det(DT(\mathbf{x}))|\mathrm{vol}(K_\delta)$. More precisely, one can prove that $\mathrm{vol}(T(K_\delta))=|\det(DT(\mathbf{x}))|\mathrm{vol}(K_\delta)+o(\delta^n)$. Note that $\mathrm{vol}(K_\delta)=c_n\delta^n$ and hence the first term is the dominant one, provided $|\det(DT(\mathbf{x})|\ne 0$. This tells us how the volume of a small set K changes under T.

(4) Again take T to be a general continuously differentiable transformation, but now we want to consider K that is not necessarily small. The idea is to break it into small pieces, apply the approximate change of volumes from the previous calculation, and then add everything up. We will not give a proof, but it is not hard to see that we should get the following:

$$\mathrm{vol}(T(K))=\int_K |JT(\mathbf{x})|d\mathbf{x}$$

where $JT(\mathbf{x})=\det(DT(\mathbf{x}))$ is the *Jacobian determinant*.

17. CHANGE OF VARIABLES IN HIGHER DIMENSIONS

Let $X=(X_1,\dots,X_n)^t$ be a random vector on a probability space with pdf f. Let $T:\mathbb{R}^n\to\mathbb{R}^n$ be a one-one function. Let $Y=(Y_1,\dots,Y_n)^t:=T(X)$. Assume that T is continuously differentiable with derivative $DT(\mathbf{x})=(\partial_i T_j(\mathbf{x}))$ and that the $DT(\mathbf{x})$ is non-singular for all $\mathbf{x}\in\mathbb{R}^n$. Then Y has a density given by

$$g(\mathbf{y})=f(T^{-1}\mathbf{y})\left|JT^{-1}(\mathbf{y})\right|$$

for $\mathbf{y}\in\mathcal{R}(T)$, where $JT^{-1}(\mathbf{y})=\det\left(DT^{-1}(\mathbf{y})\right)$ is the *Jacobian determinant* of T^{-1}.

Heuristic reason: The meaning of saying that the random vector X has density f means that for any $\mathbf{x}$ and for small δ we have $\mathbf{P}\{X\in B(\mathbf{x},\delta)\}\approx f(\mathbf{x})\mathrm{vol}(B(\mathbf{x},\delta))$.

Similarly, if g is the density of $Y=T(X)$ at a point $\mathbf{y}$, then $\mathbf{P}\{Y\in B(\mathbf{y},\delta)\}\approx g(\mathbf{y})\mathrm{vol}(B(\mathbf{y},\delta))$.

On the other hand, because T is one-one,

$$\mathbf{P}\{Y\in B(\mathbf{y},\delta)\}=\mathbf{P}\{X\in T^{-1}(B(\mathbf{y},\delta)\}\approx f(T^{-1}\mathbf{x})\mathrm{vol}(T^{-1}(B(\mathbf{y},\delta)).$$

By the volume change discussion in the previous section, and because T^{-1} is assumed to be continuously differentiable, we can write $\text{vol}(T^{-1}(B(\mathbf{y},\delta)) \approx |JT^{-1}(\mathbf{y})|\text{vol}(B(\mathbf{y},\delta))$.

Putting everything together we get $g(\mathbf{y}) = f(T^{-1}\mathbf{y})|JT^{-1}(\mathbf{y})|$.

Example 65. Let X have pdf f on $\mathbb{R}^n$. Let $T(\mathbf{x}) = A\mathbf{x}+\mathbf{b}$ where A is a non-singular $n\times n$ matrix and $\mathbf{b}\in\mathbb{R}^n$. Then, $Y = T(X) = AX+\mathbf{b}$. As A is non-singular, $T^{-1}\mathbf{y} = A^{-1}(\mathbf{y}-\mathbf{b})$ and $DT^{-1}(\mathbf{y}) = A^{-1}$. Thus, Y has pdf

$$g(\mathbf{y}) = f(A^{-1}(\mathbf{y}-\mathbf{b}))|\det(A^{-1})|.$$

Take the case of $X \sim N_n(0,I)$. Then $f(\mathbf{x}) = (2\pi)^{-n/2}\exp\{-\frac{1}{2}\mathbf{x}^t\mathbf{x}\}$. Hence we get

$$g(\mathbf{y}) = (2\pi)^{-n/2}|\det(A^{-1})|\exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{b})^t(AA^t)^{-1}(\mathbf{y}-\mathbf{b})\} = (2\pi)^{-n/2}\sqrt{\det(\Sigma)}\exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{b})^t\Sigma^{-1}(\mathbf{y}-\mathbf{b})\}.$$

where $\Sigma \sim AA^t$. Thus, $Y \sim N_n(\mathbf{b},AA^t)$.

Example 66. Let (X_1,X_2) have density $f(x_1,x_2) = \lambda^2 e^{-\lambda(x_1+x_2)}\mathbf{1}_{x_1>0}\mathbf{1}_{x_2>0}$. Then, it is easy to see that X_1 and X_2 are both $\text{Exp}(\lambda)$ random variables.

Define $T(x_1,x_2) = (x_1+x_2, \frac{x_1}{x_1+x_2})$. This maps $\mathbb{R}_+^2$ onto $\mathbb{R}_+\times(0,1)$ (since X has range $\mathbb{R}_+^2$, it is enough for T to be defined on this subset of $\mathbb{R}^2$) with inverse $T^{-1}(u,v) = (uv,u(1-v))$. Thus $DT^{-1}(u,v) = \begin{bmatrix} v & u \\ 1-v & -u \end{bmatrix}$ and $JT^{-1}(u,v) = -u$. Therefore, $U := X_1+X_2$ and $V = X_1/(X_1+X_2)$ has joint density

$$g(u,v) = f(T^{-1}(u,v))|JT^{-1}(u,v)| = \lambda^2 e^{-\lambda(uv+u(1-v))}u\mathbf{1}_{u>0}\mathbf{1}_{v\in(0,1)} = \lambda^2 ue^{-\lambda u}\mathbf{1}_{u>0}\mathbf{1}_{v\in(0,1)}.$$

This is of the form $g_1(u)g_2(v)$ where $g_1(u) = \lambda^2 ue^{-\lambda u}\mathbf{1}_{u>0}$ and $g_2(v) = \mathbf{1}_{v\in(0,1)}$. Thus, $U \sim \text{Gamma}(2,\lambda)$ while $V \sim \text{unif}[0,1]$.

Example 67. Let $X = (X_1,X_2)$ have density $f(\mathbf{x}) = \lambda^2\exp\{-\lambda(x_1+x_2)\}\mathbf{1}_{x_1>0}\mathbf{1}_{x_2>0}$. Let $T(X) = (x_1+x_2, \frac{x_1}{x_1+x_2})$. Then T is one-one on $\mathbb{R}_+^2$ and maps it onto $\mathbb{R}_+\times(0,1)$. Clearly, $T^{-1}(\mathbf{y}) = (y_1y_2, y_1(1-y_2))$ and $DT^{-1}(\mathbf{y}) = \begin{bmatrix} y_2 & y_1 \\ 1-y_2 & -y_1 \end{bmatrix}$ and hence $JT^{-1}(\mathbf{y}) = -y_1$. Hence, $Y = T(X)$ has density

$$g(\mathbf{y}) = \lambda^2 y_1\exp\{-\lambda y_1\}\mathbf{1}_{y_1>0}\mathbf{1}_{y_2\in(0,1)}.$$

From this we can get the marginal densities of $Y_1 = X_1+X_2$ and $Y_2 = \frac{X_1}{X_1+X_2}$ as

$$h_1(y_1) = \int_{\mathbb{R}} g(y_1,y_2)dy_2 = \lambda^2 y_1\exp\{-\lambda y_1\}\mathbf{1}_{y_1>0}, \qquad h_2(y_2) = \int_{\mathbb{R}} g(y_1,y_2)dy_1 = \mathbf{1}_{y_2\in(0,1)}$$

which shows that $Y_1 \sim \text{Gamma}(2,\lambda)$ (as we already saw in the previous example) and Y_2 has unif([0,1]) distribution.

Example 68. Let $X = (X_1,X_2) \sim N_2(0,I)$. Let $T(\mathbf{x}) = (x_1/x_2, x_2)$. T is well-defined only on the set $S = \{\mathbf{x} : x_2 \neq 0\}$ but since $\mathbf{P}\{X\in S\} = 1$, we can still apply the change of variable formula. Note that on $T : S \to S$ is one-one and onto with $T^{-1}(\mathbf{y}) = (y_1y_2, y_2)$. Hence, $DT^{-1}(\mathbf{y}) = \begin{bmatrix} y_2 & y_1 \\ 0 & 1 \end{bmatrix}$ and $JT^{-1}(\mathbf{y}) = y_2$. Thus $Y = (X_1/X_2, X_2)$ has pdf

$$g(\mathbf{y}) = f(y_1y_2,y_2)|y_2| = \frac{|y_2|}{2\pi}\exp\left\{-\frac{y_2^2(1+y_1^2)}{2}\right\}.$$

From this we can get the density of X_1/X_2 to be

$$h(y_1) = \int_{\mathbb{R}} g(y_1,y_2)dy_2 = \frac{2}{2\pi}\int_0^\infty y_2\exp\left\{-\frac{y_2^2(1+y_1^2)}{2}\right\}dy_2 = \frac{1}{\pi}\int_0^\infty \exp\left\{-u(1+y_1^2)\right\}du = \frac{1}{\pi(1+y_1^2)}.$$

Thus X_1/X_2 has standard Cauchy distribution.