Probability Integral Transform and change of variables

Suppose a random variable X has :

probability density function	=	$p_X(x)$
cumulative distribution function	=	$P_X(x)$

then its probability integral transform is defined as :

y	=	$P_X(x)$	probability integral transform
x	=	$P_X^{-1}(y)$	probability integral inverse transform

which is equivalent to the cumulative distribution function of X. There are two properties for the probability integral transform. (1) The output after probability integral transform, i.e. y, is a random variable with uniform distribution in range [0, 1], in other words, cumulative distribution function is a transformation of a random variable into a new random variable with uniform distibution. (2) The inverse of probability integral transform of a uniform distributed random variable gives a random variable with probability density function p(x), which is the first order derivative of the probability integral transform, hence we can generate random variable by making use of this property.

Property 1

If $\quad y = P_X(x) = \int_{-\infty}^{x} p(z)dz \qquad \forall y \in [0,1]$

Then $\quad y \sim uniform(0,1)$

proof

$$\begin{aligned}
p_Y(y) &= \frac{d}{dy}\Pr ob(Y > y) && \forall y \in [0,1] \\
&= \frac{d}{dy}\Pr ob(X > x) && \forall y \in [0,1] \\
&= \frac{dx}{dy}\frac{dP_X(x)}{dx} && \forall y \in [0,1] \\
&= \frac{dx}{dy}\frac{dy}{dx} && \forall y \in [0,1] \\
&= 1
\end{aligned}$$

Property 2

If $\quad y \sim uniform(0,1)$

and $\quad x = f^{-1}(y) \qquad \forall y \in [0,1]$

then $\quad p_X(x) = \frac{df(x)}{dx}$

proof

$$\begin{aligned}
p_X(x) &= \frac{d}{dx}\Pr ob(X > x) \\
&= \frac{d}{dx}\Pr ob(Y > y = f(x)) && \forall y \in [0,1] \\
&= \frac{dy}{dx}\frac{d}{dy}\Pr ob(Y > y = f(x)) && \forall y \in [0,1] \\
&= \frac{dy}{dx}uniform(0,1) && \forall y \in [0,1] \\
&= \frac{df(x)}{dx}
\end{aligned}$$

Change of variable in 1D density function

Recall the relation between probrobability, cumulative distribution function *Px(x)* and probability density function *px(x)* :

$$
\begin{aligned}
prob(X \le x) &= P_X(x) \\
prob(x < X \le x+dx) &= P_X(x+dx) - P_X(x) \\
&= dP_X(x) && \textit{lets expand this differential by first term in Taylor series} \\
&= p_X(x)dx && \textit{since } dP_X(x)/dx = p_X(x)
\end{aligned}
$$

$$
\boxed{prob(x < X \le x+dx) = p_X(x)dx \qquad \textit{equation 1}}
$$

Given the *1D* probability density in *X*, what will be the *1D* probability density in *Y* when we apply change of variable $X = f(Y)$?

$$
prob(y < Y \le y+dy) = prob(x < X \le x+dx)
$$

This is true only if the point *x* corresponds to the point *y*, and the point *x+dx* corresponds to the point *y+dy*, i.e.

$$
\begin{aligned}
x &= f(y) && \textit{equation *} \\
\text{and} \quad x+dx &= f(y+dy) = f(y) + \partial_y f(y)dy && \textit{equation **} \\
\text{i.e.} \quad dx &= \partial_y f(y)dy && \textit{applying equation * to equation **}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \quad prob(y < Y \le y+dy) &= prob(x < X \le x+dx) \\
p_Y(y)dy &= p_X(x = f(y))dx && \textit{applying equation 1 to both LHS and RHS} \\
&= p_X(x = f(y))\partial_y f(y)dy \\
p_Y(y) &= p_X(x = f(y))\partial_y f(y)
\end{aligned}
$$

Change of variables in 2D density function

Given the *2D* probability density in *(X,Y)*, what will be the *2D* probability density in *(U,V)* when we apply change of variables $X = f(U,V)$ and $Y = g(U,V)$?

$$prob(u < U \le u+du, v < V \le v+dv) \quad = \quad prob(x < X \le x+dx, y < Y \le y+dy) \qquad \textit{equation 2}$$

This is true only if : (1) *(x,y)* and *(u,v)* refer to the same point and
(2) *(x+dx,y+dy)* and *(u+du,v+dv)* refer to the same point, which imply :

$$\begin{array}{llll} & x = f(u,v) & & \\ \text{and} & y = g(u,v) & & \\ \text{and} & dx = \partial_u f du + \partial_v f dv & \textit{by Taylor series} & \\ \text{and} & dy = \partial_u g du + \partial_v g dv & \textit{by Taylor series} & \textit{equation 3} \end{array}$$

When we apply equation *1* (*2D* version) to both sides of equation *2*, we have :

$$\begin{array}{lll} p_{U,V}(u,v)dudv & = p_{X,Y}(x,y)dxdy & \\ & = p_{X,Y}(x=f(u,v),y=g(u,v))(\partial_u f du + \partial_v f dv)(\partial_u g du + \partial_v g dv) & \textit{applying equation 3} \\ & = p_{X,Y}(x=f(u,v),y=g(u,v))(\partial_u f \partial_u g(du)^2 + \partial_u f \partial_v g(dudv) + \partial_v f \partial_u g(dudv) + \partial_v f \partial_v g(dv)^2) & \\ & = p_{X,Y}(x=f(u,v),y=g(u,v))(\partial_u f \partial_v g + \partial_v f \partial_u g)dudv & (du)^2 = 0 \text{ and } (dv)^2 = 0 \\ p_{U,V}(u,v) & = p_{X,Y}(x=f(u,v),y=g(u,v))(\partial_u f \partial_v g + \partial_v f \partial_u g) & \\ & = p_{X,Y}(x=f(u,v),y=g(u,v))\det(J) & \textit{equation 4} \end{array}$$

$$\text{where} \qquad J = \begin{bmatrix} \partial_u f & \partial_v f \\ \partial_u g & \partial_v g \end{bmatrix}$$

Equation 4 provides a simple method to obtain density function in new domain after changing of variables. To summarise, suppose area *A* in domain *(x,y)* and area *B* in domain *(u,v)* refer to the same piece of integrating region, then :

$$\begin{array}{llll} & p_{U,V}(u,v) & = p_{X,Y}(x,y)\det(J) & = p_{X,Y}(x=f(u,v),y=g(u,v))\det(J) \\ & p_{U,V}(u,v)dudv & = p_{X,Y}(x,y)\det(J)dudv & = p_{X,Y}(x=f(u,v),y=g(u,v))\det(J)dudv \\ & \iint_B p_{U,V}(u,v)dudv & = \iint_B p_{X,Y}(x,y)\det(J)dudv & = \iint_B p_{X,Y}(x=f(u,v),y=g(u,v))\det(J)dudv \\ \text{and} & \iint_B p_{U,V}(u,v)dudv & = \iint_A p_{X,Y}(x,y)dxdy & \\ \text{but} & p_{U,V}(u,v)dudv & \ne p_{X,Y}(x,y)dxdy & \\ & dudv & \ne \det(J)^{-1}dxdy & \end{array}$$

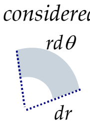
For example, if *(u,v)* is the polar coodinate :

$$\begin{array}{llll} x & = r\cos(\theta) & & \\ y & = r\sin(\theta) & & \\ J & = \begin{bmatrix} \partial_r f & \partial_\theta f \\ \partial_r g & \partial_\theta g \end{bmatrix} & = \begin{bmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{bmatrix} & \\ |J| & = r\cos^2(\theta) + r\sin^2(\theta) & = r & \textit{equation 5} \end{array}$$

Hence we have :

$$\begin{array}{llll} & p_{R,T}(r,\theta) & = p_{X,Y}(x,y)r & = p_{X,Y}(x=r\cos(\theta),y=r\sin(\theta))r \\ & p_{R,T}(r,\theta)drd\theta & = p_{X,Y}(x,y)rdrd\theta & = p_{X,Y}(x=r\cos(\theta),y=r\sin(\theta))rdrd\theta \\ & \int_0^{2\pi}\int_0^{\infty} p_{R,T}(r,\theta)drd\theta & = \int_0^{2\pi}\int_0^{\infty} p_{X,Y}(x,y)rdrd\theta & = p_{X,Y}(x=r\cos(\theta),y=r\sin(\theta))rdrd\theta \\ \text{and} & \int_0^{2\pi}\int_0^{\infty} p_{R,T}(r,\theta)drd\theta & = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y}(x,y)dxdy & \\ \text{but} & p_{R,T}(r,\theta)drd\theta & \ne p_{X,Y}(x,y)dxdy & \\ & drd\theta & \ne r^{-1}dxdy & \\ \text{i.e.} & dxdy & \ne rdrd\theta & \end{array}$$

$rdrd\theta$ can be considered as arc area



Box Muller algorithm
Standard normal x can be generated by uniform random number y using probability integral inverse transform, however we cant get the close form solution for inverse f.

$$
\begin{aligned}
& x &=& \; f^{-1}(y) \\
\text{where} \quad & y &=& \; f(x) &=& \; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}dz
\end{aligned}
$$

Box Muller algorithm provides a solution. Suppose x and y are two independent normal random variables. Lets make a change of variables as $x = r\cos\theta$ and $y = r\sin\theta$. The joint probability density function in (r,θ) space is :

$$
\begin{aligned}
p_{R,T}(r,\theta) &= p_{X,Y}(x,y)\det(J) && \textit{applying equation 4} \\
&= p_X(x = r\cos(\theta))\, p_Y(y = r\sin(\theta))\det(J) && \textit{x \& y are independent} \\
&= \frac{1}{\sqrt{2\pi}}e^{-(r\cos\theta)^2/2}\frac{1}{\sqrt{2\pi}}e^{-(r\sin\theta)^2/2} \times r && \textit{applying equation 5} \\
&= \frac{1}{2\pi}e^{-r^2/2} \times r
\end{aligned}
$$

$$
\begin{aligned}
p_R(r) &= \int_0^{2\pi} p_{R,T}(r,\theta)d\theta &=& \; e^{-r^2/2}r && \textit{integrate out } \theta \\
p_T(\theta) &= \int_0^{\infty} p_{R,T}(r,\theta)dr &=& \; \frac{1}{2\pi}\int_0^{\infty} e^{-r^2/2}rdr && \textit{integrate out } r \\
& &=& \; \frac{1}{2\pi}\int_0^{\infty} e^{-r^2/2}d(r^2/2) \\
& &=& \; \frac{-1}{2\pi}[e^{-r^2/2}]_0^{\infty} \\
& &=& \; \frac{1}{2\pi}1_{\theta\in[0,2\pi]} && \textit{uniform within } [0,2\pi]\textit{, independent of } \theta
\end{aligned}
$$

Hence r is a random variable with pdf $p_R(r)$, while θ is a uniform random variable. Therefore if we can generate r and θ separately, then we can obtain independent standard normal x and y by applying $x = r\cos\theta$ and $y = r\sin\theta$. The generation θ is straight forward, as it is uniform within $[0,2\pi]$, while the generation of r can be achieved by applying probability integral inverse transform on $P_R(r)$.

$$
\begin{aligned}
P_R(r) &= \int_0^r p_R(z)dz \\
&= \int_0^r e^{-z^2/2}zdz \\
&= \int_0^r e^{-z^2/2}d(z^2/2) \\
&= [-e^{-z^2/2}]_0^r \\
&= 1 - e^{-r^2/2}
\end{aligned}
$$

Suppose u and v are two independent *1D* uniform random numbers *[0,1]* :

$$
\begin{aligned}
u &= 1 - e^{-r^2/2} \\
r &= \sqrt{-2\ln(1-u)} \;=\; \sqrt{-2\ln u} \\
\theta &= 2\pi v
\end{aligned}
$$

Here is the solution.

$$
\begin{aligned}
x &= r\cos\theta &=& \; \sqrt{-2\ln u}\cos(2\pi v) \\
y &= r\sin\theta &=& \; \sqrt{-2\ln u}\sin(2\pi v)
\end{aligned}
$$

space (x,y)	$x = r\cos\theta$	*space* (r,θ)	*inv transform*	*space (u,v)*
x ~ *normal*	←	$r \sim p_R(r)$	←	u ~ *uniform(0,1)*
y ~ *normal*	←	θ ~ *uniform*$(0,2\pi)$	←	v ~ *uniform(0,1)*
	$y = r\sin\theta$		*scaling*	

Marsaglia Polar Method

However the *sin* and *cos* operations in Box Muller algorithm are time consuming, we can improve it using Marsaglia polar method. Consider space *(u',v')* with uniform joint density within the unit circle, i.e.

$$p_{U',V'}(u',v') \quad = \quad (1/\pi)\times 1_{u'^2+v'^2\le 1}$$

Please note that scale $1/\pi$ is the normalising factor for density function to ensure its integration equals to *1* (recall : area of unit circle equals to π). It can be implemented by : `u'=2*rand(0,1)-1` and `v'=2*rand(0,1)-1` which should be repeated until `u'^2+v'^2<=1`. Now lets look step by step how *(u',v')* can helps to generate *(x,y)*. Now make a change of variables from *(u',v')* to *(s,φ)*.

$$\begin{aligned} u' &= \sqrt{s}\cos\varphi \\ \text{and} \quad v' &= \sqrt{s}\sin\varphi \end{aligned} \qquad \textbf{or} \qquad \begin{aligned} s &= u'^2+v'^2 \\ \text{and} \quad \varphi &= \tan^{-1}(v'/u') \end{aligned}$$

$$\begin{aligned}
p_{S,\Phi}(s,\varphi) &= p_{U',V'}(u=\sqrt{s}\cos\varphi, v=\sqrt{s}\sin\varphi)\det(J) \\
&= \frac{1}{\pi}1_{u'^2+v'^2\le 1}\times\det\begin{bmatrix}\partial_s\sqrt{s}\cos\varphi & \partial_\varphi\sqrt{s}\cos\varphi \\ \partial_s\sqrt{s}\sin\varphi & \partial_\varphi\sqrt{s}\sin\varphi\end{bmatrix} \\
&= \frac{1}{\pi}1_{s\le 1}\times\det\begin{bmatrix}(2\sqrt{s})^{-1}\cos\varphi & -\sqrt{s}\sin\varphi \\ (2\sqrt{s})^{-1}\sin\varphi & +\sqrt{s}\cos\varphi\end{bmatrix} \\
&= \frac{1}{\pi}1_{s\le 1}\times\frac{1}{2}(\cos^2\varphi+\sin^2\varphi) \\
&= \frac{1}{2\pi}1_{s\le 1} \\
p_S(s) &= \int_0^{2\pi}p_{S,\Phi}(s,\varphi)d\varphi = \int_0^{2\pi}\frac{1}{2\pi}1_{s\le 1}d\varphi = 1_{s\le 1} \\
p_\Phi(\varphi) &= \int_0^{\infty}p_{S,\Phi}(s,\varphi)d\varphi = \int_0^{\infty}\frac{1}{2\pi}1_{s\le 1}ds = \frac{1}{2\pi}
\end{aligned}$$

Thus :

$$\begin{aligned}
s &= u'^2+v'^2 && \sim uniform(0,1) \\
\varphi &= \tan^{-1}(v'/u') && \sim uniform(0,2\pi)/2\pi \\
\Rightarrow \quad \cos\varphi &= \cos(\tan^{-1}(v'/u')) && = u'/\sqrt{s} \\
\sin\varphi &= \cos(\tan^{-1}(v'/u')) && = v'/\sqrt{s}
\end{aligned}$$

From Box Muller, we can derive the Marsaglia polar method :

$$\begin{aligned}
x &= \sqrt{-2\ln u}\cos(2\pi v) & y &= \sqrt{-2\ln u}\sin(2\pi v) & & \\
&= \sqrt{-2\ln s}\cos(2\pi v) & &= \sqrt{-2\ln s}\sin(2\pi v) & &\text{since } u\sim uniform(0,1) \text{ and } s\sim uniform(0,1) \\
&= \sqrt{-2\ln s}\cos(\varphi) & &= \sqrt{-2\ln s}\sin(\varphi) & &\text{since } v\sim uniform(0,1) \text{ and } \varphi\sim uniform(0,2\pi)/2\pi \\
&= \sqrt{-2\ln s}(u'/\sqrt{s}) & &= \sqrt{-2\ln s}(v'/\sqrt{s}) & &\text{since } \cos\varphi\sim u'/\sqrt{s} \text{ and } \sin\varphi\sim v'/\sqrt{s}
\end{aligned}$$

Unlike (u,v) in Box Muller, (u',v') in Marsaglia polar method is uniform over unit circle : *(-1,+1)*