Random Variate Generation from Continuous Distribution Function

Shim, Hyung Jin

Nuclear Engineering Department,
Seoul National University

References

- Reuven Y. Rubinstein, "Simulation and the Monte Carlo Method," John Wiley & Sons, Inc., New York (1981).
- C. J. Everett and E. D. Cashwell, "A Third Monte Carlo Sampler" Los Alamos report, LA-9721-MS (1983).
- H. J. Shim, “McCARD Methodology,” SNUMCL/TR001/2010(01), 2012.

3.1.2.1 RVG – Principles (Inverse Transform & Rejection Methods)

Inverse Transform Method

- From a probability density function (PDF), $f(x)$ for $a \le x \le b$, the corresponding cumulative probability density function (CDF), $F(x)$ can be defined as

$$F(x) = \int_a^x f(x')dx'$$

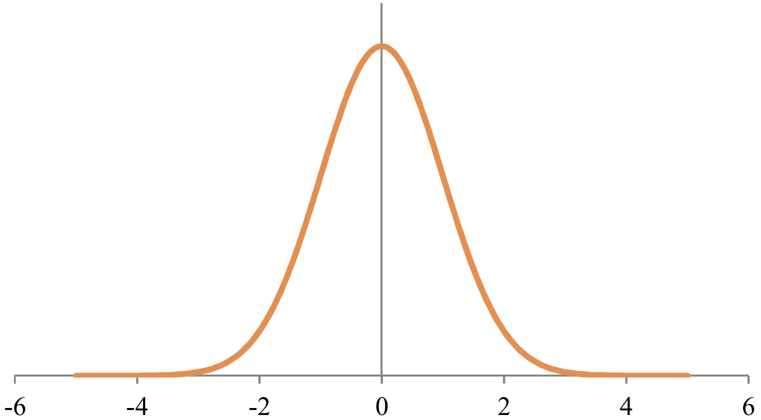
- When a random variable X follows a PDF, $f(x)$ and its corresponding CDF, $F(x)$, it can be sampled using a random number, ξ, which is sampled from a uniform distribution in interval (0,1), as

$$\boxed{\xi = F(x) \ \Rightarrow \ X = F^{-1}(\xi)}$$

- Proof:

$$P(X \le x) = P\left[F^{-1}(\xi) \le x\right] = P\left[\xi \le F(x)\right] = F(x)$$

Explanatory Diagram of Inverse Transform Method



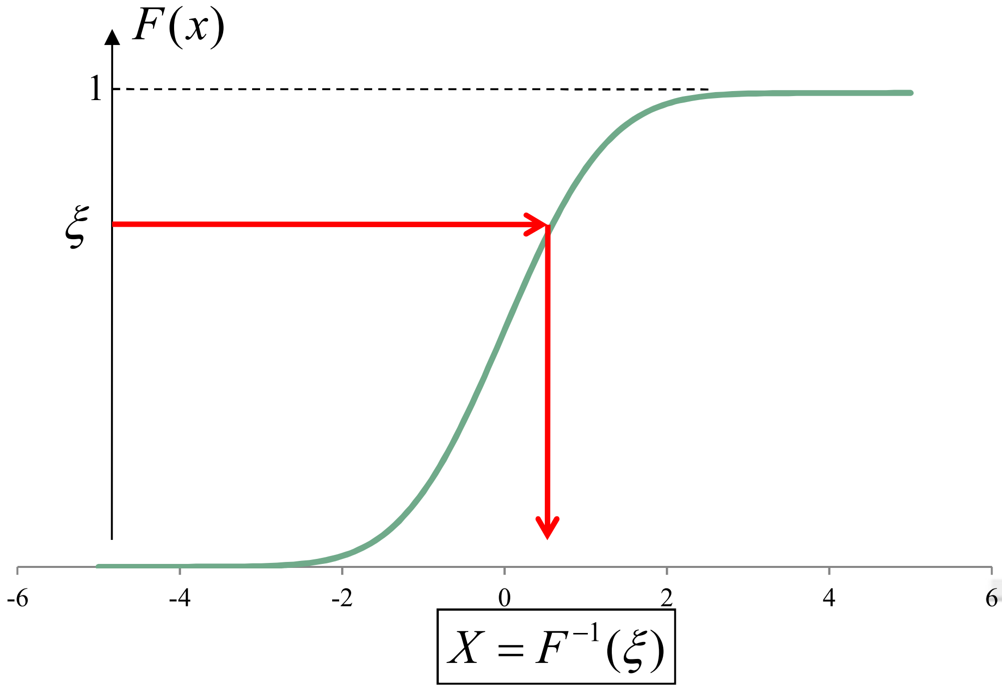
< Normal Distribution >

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right);$$

$$\sigma = 1,\ \ \mu = 0$$

$$F(x) = \frac{1}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right);$$

$$erf(X) = \frac{2}{\sqrt{\pi}} \int_0^X e^{-t^2}\, dt$$

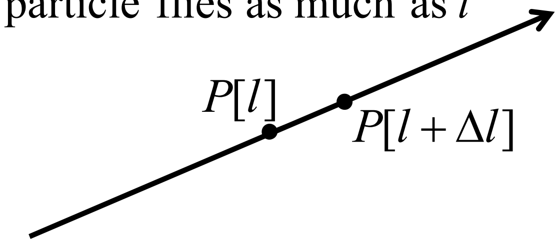


Example #1 – Sampling the flight length

- A probability that a particle flies as long as ℓ and collides with a atom can be written as

$$p(l) = \Sigma_t \underline{\exp(-\Sigma_t l)}$$

$P[l]$ = probability that the particle flies as much as l



$$P(l) - P(l+\Delta l) = P(l)\cdot\Sigma_t \Delta l$$

$$\Rightarrow P'(l) = -\Sigma_t P(l)$$

- Then, the flight length can be sampled by

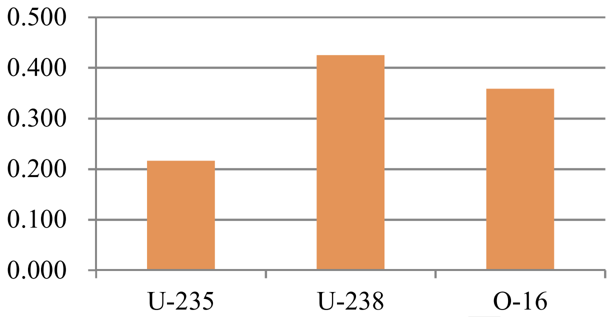
$$\xi = \int_0^x \Sigma_t \exp(-\Sigma_t l)dl$$

⬇

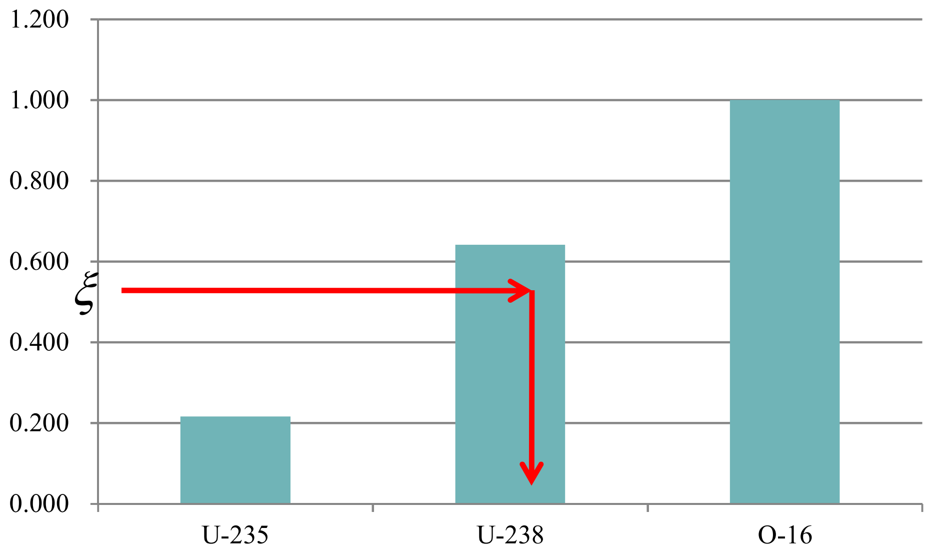
$$x = -\frac{\ln(1-\xi)}{\Sigma_t} = -\frac{\ln\xi'}{\Sigma_t}$$

Example #2 – Selection of a Collided Nuclide

Nuclide	Σ_t	PDF	CDF
U-235	0.107	0.216	0.216
U-238	0.211	0.425	0.641
O-16	0.178	0.359	1.000



$$\sum_{i=1}^{k-1}\Sigma_t^i < \xi \sum_{i=1}^{N}\Sigma_t^i \le \sum_{i=1}^{k}\Sigma_t^i$$

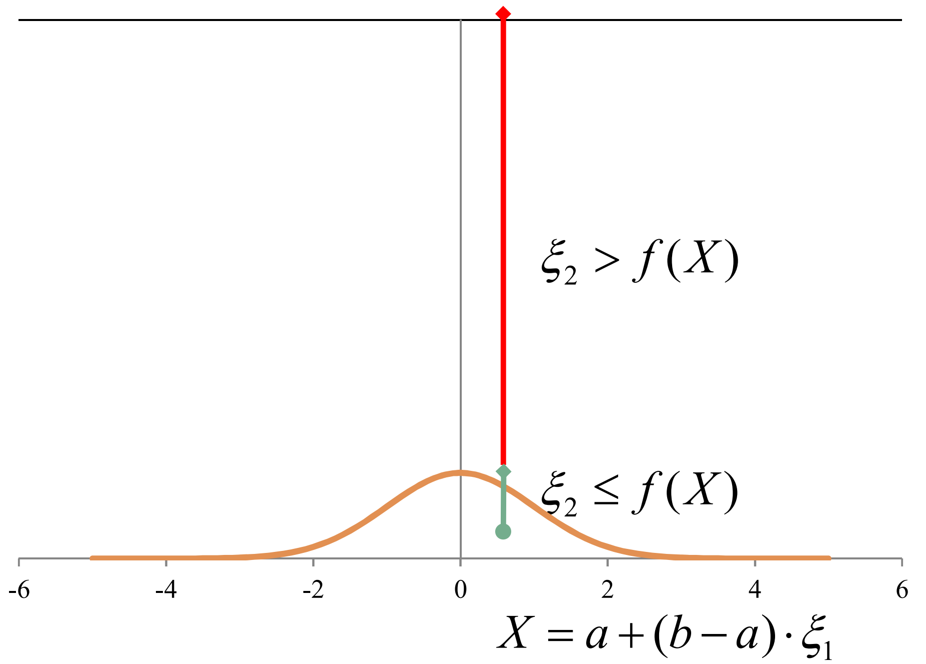


Acceptance – Rejection Method – 1/2

- It is common that the CDF and its inverse function for a random variable cannot be analytically obtained.
- A random variable *X*, which follows the PDF, *f*(*x*) in interval [a,b] can be sampled by trial and error as

  ① Sample *X* by $X = a + (b-a)\cdot\xi_1$ using a random number ξ_1.

  ② From another random number ξ_2, accept *X* if $\xi_2 \le f(X)$ and return to ① elsewhere.



Acceptance – Rejection Method – 2/2

- In order to enhance the sampling efficiency, the PDF f(x) can be represented as

$$f(x) = Ch(x)g(x)$$

where $C \geq 1$, $h(x)$ is also a PDF, and $0 < g(x) \leq 1$.

- Then X can be sampled as

① Sample X from the PDF of h(x).

② Using a random number ξ, accept X if $\xi \leq g(X)$ and reject elsewhere.

Example #1 – Sampling from Normal Distribution

- The standard normal distribution can be expressed as

$$f(x) = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) = Ch(x)g(x), \;\; x \geq 0;$$

$$C = \sqrt{\frac{2e}{\pi}}, \;\; h(x) = e^{-x}, \;\; g(x) = \exp\left[-\frac{(x-1)^2}{2}\right]$$

- Then X can be sampled by

① Sample X from h(x).

$$\int_0^x e^{-x'} dx' = 1 - e^{-x} = \xi \Rightarrow X = -\ln(1-\xi) \Rightarrow X = -\ln \xi$$

② If the below condition is satisfies, accept X. The condition is violated, go to step ①.

$$\xi \leq \exp\left[-\frac{(X-1)^2}{2}\right] \Rightarrow -\ln \xi \leq \frac{(X-1)^2}{2}$$

Example #2 – Sampling from Isotropic Distribution

- Method 1

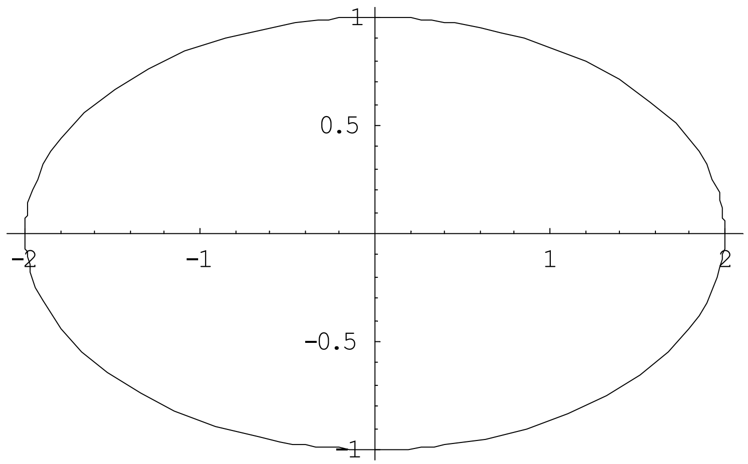
```
phi=2*PI*RNG->GetRN();
sinP=sin(phi); cosP=cos(phi);
```

- Method 2

```
do {
   C1=2.*RNG->GetRN()-1.;
   C2=2.*RNG->GetRN()-1.;
   C3=C1*C1+C2*C2;
}while(C3>1.);
C4=sqrt(C3); sinP=C1/C4; cosP=C2/C4;
```

3.1.2.2 Substitution of Variables in Multiple Integration

Compute the Area of an Ellipse



$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

$$A = \iint_{\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 \le 1} dxdy = ?$$

- By substituting x and y as

$$u = \frac{x}{a}, \quad v = \frac{y}{b} \quad \Rightarrow \quad dx = adu, \quad dy = bdv$$

the double integrals can be computed by

$$A = \iint_{\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 \le 1} dxdy$$

$$= \iint_{u^2 + v^2 \le 1} abdudv = \pi ab$$

Integration by Linear Transformation

- In the double integrations, we are going to change the original set of variables (x,y) to a set of other variables (u,v).

$$\int\int \cdots dxdy = \int\int \cdots dudv$$

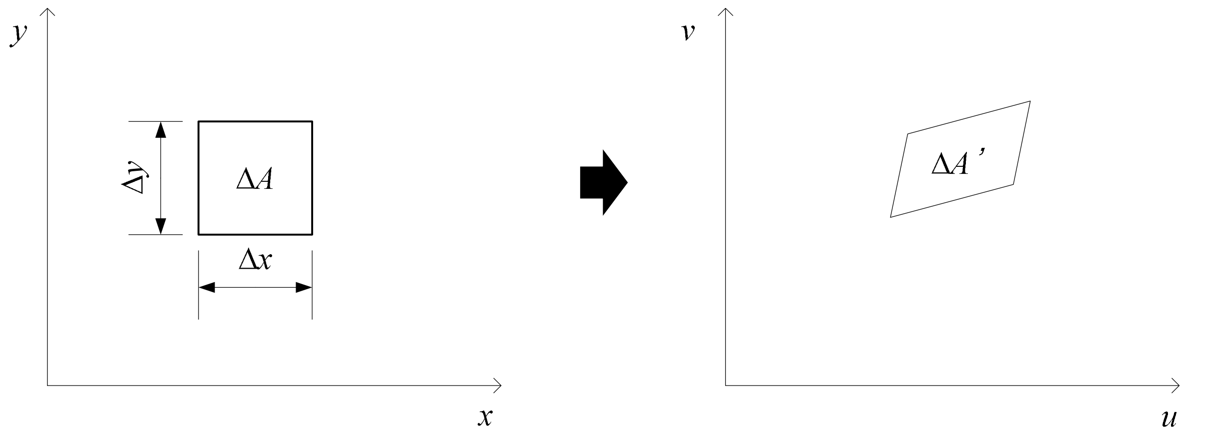
- Then our problem becomes how to convert a small area of ΔA=$\Delta x \Delta y$ to the corresponding area of $\Delta A'$=$\Delta u \Delta v$.

- Consider the two-dimensional linear transformation as

$$u = a_{11}x + a_{12}y,$$

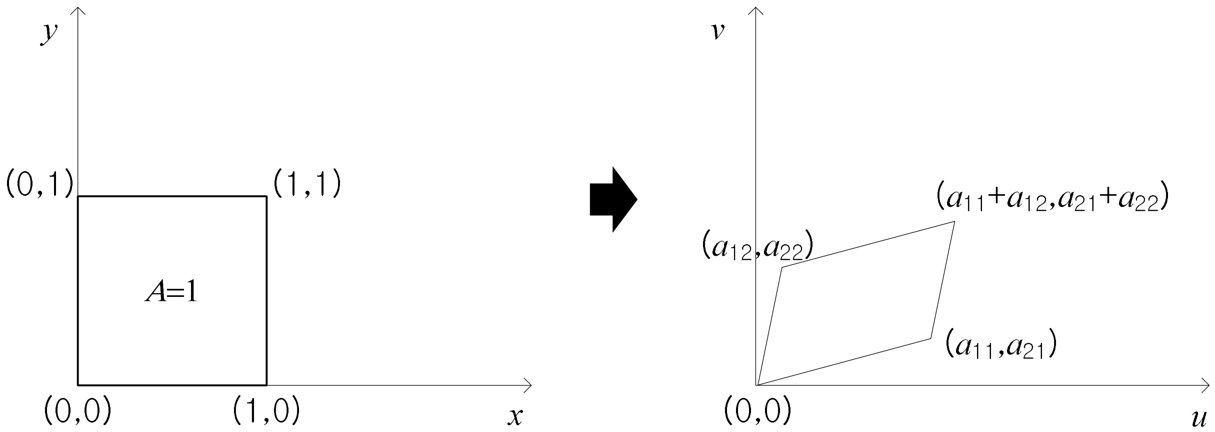
$$v = a_{21}x + a_{22}y$$

- Then the rectangular area is converted to the area of a parallelogram as



Integration by Linear Transformation (Contd.)

- From the property of the linear transformation, the scaling factor F in $\Delta A' = F\Delta A$ is independent of its location and size.



- The scaling factor F becomes the area of the parallelogram:

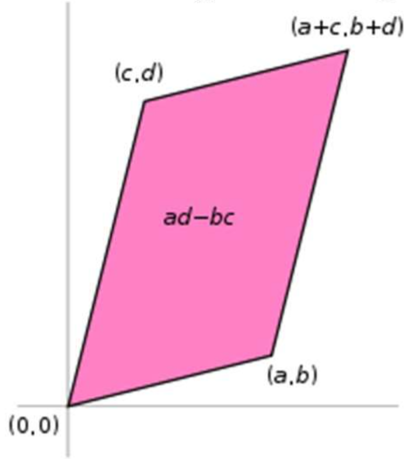
$$F = A' = a_{11}a_{22} - a_{12}a_{21}$$

- Therefore the integration becomes

$$\int\int \cdots dxdy = \int\int \cdots F^{-1} dudv$$

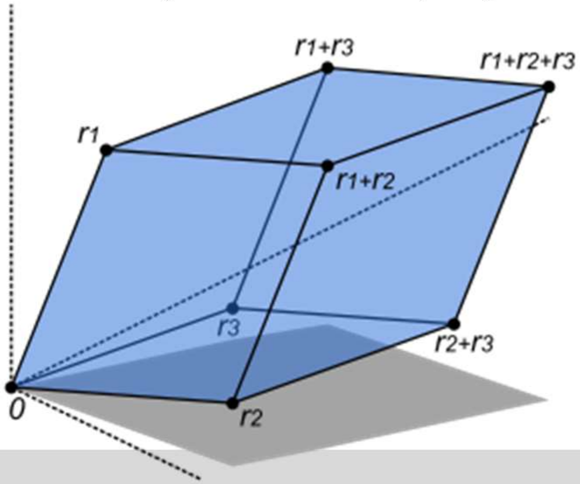
Meaning of Determinant

- The area of the parallelogram is the absolute value of the determinant of the matrix formed by the vectors representing the parallelogram's sides.



$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- The volume of the parallelepiped is the absolute value of the determinant of the matrix formed by the rows $r1$, $r2$, and $r3$.



$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Properties of Determinant

$$\det(A^T) = \det(A)$$

$$\det(A^{-1}) = 1/\det(A)$$

$$\det(AB) = \det(A)\det(B)$$

General Transformation

- When the variable x and y are transformed to u and v as

$$u \equiv u(x, y),$$
$$v \equiv v(x, y)$$

- Then, by the linear approximation Δx and Δy can be written as

$$\Delta u \cong u_x \Delta x + u_y \Delta y, \;\; \Delta v \cong v_x \Delta x + v_y \Delta y;$$

$$f_\alpha = \frac{\partial f}{\partial \alpha} \;\; (f = u \text{ or } v, \;\; \alpha = x \text{ or } y)$$

$$\Rightarrow \quad \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

- In the same way to the linear transformation, the small rectangular area in x-y coordinate becomes a small area of the corresponding parallelogram in u-v coordinate.

$$(\Delta x, 0) \Rightarrow (u_x \Delta x, v_x \Delta x)$$
$$(0, \Delta y) \Rightarrow (u_y \Delta y, v_y \Delta y)$$

$$\Rightarrow \quad Area' = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} \Delta x \Delta y$$

Jacobian Matrix

- In vector calculus, the Jacobian matrix is the matrix of all first-order partial derivatives of a vector- or scalar-valued function with respect to another vector.

$$J = \begin{pmatrix} \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial x_1} & \cdots & \dfrac{\partial F_m}{\partial x_n} \end{pmatrix} \equiv \frac{\partial(F_1,\cdots,\partial F_m)}{\partial(x_1,\cdots,x_n)}$$

- According to the inverse function theorem, the matrix inverse of the Jacobian matrix of an invertible function is the Jacobian matrix of the *inverse* function.

$$J(\mathbf{F}^{-1}(p)) = \left[J(\mathbf{F}(p))\right]^{-1}$$

- If m=n, then **F** is a function from n-space to n-space and the Jacobian matrix is a square matrix. We can then form its determinant, known as the **Jacobian determinant**. <u>The Jacobian determinant is sometimes simply called "the Jacobian."</u>

$$J = \begin{vmatrix} \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial x_1} & \cdots & \dfrac{\partial F_m}{\partial x_n} \end{vmatrix} \equiv \frac{\partial(F_1,\cdots,\partial F_m)}{\partial(x_1,\cdots,x_n)}$$

Change of Variables in Double Integrals

- Sometimes, it is often advantageous to evaluate $\iint_R f(x,y)dxdy$ in a coordinate system other than the *xy*-coordinate system.
- The formula for change of variables is given by

$$\boxed{\iint_R f(x,y)dxdy = \iint_S f\left[x(u,v), y(u,v)\right]\left|\frac{\partial(x,y)}{\partial(u,v)}\right| dudv}$$

where |…| means the absolute value.

Example 1 of Transformation to Polar Coordinates

- Let R be the disc of radius 2 centered at the origin. Calculate

$$\iint_R \sin(x^2 + y^2)\,dxdy = ?$$

- Solution:

$$\iint_R \sin(x^2 + y^2)\,dxdy = \iint_S \sin r \left\| \begin{matrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{matrix} \right\| dr d\theta$$

$$\boxed{x = r\cos\theta, \quad y = r\sin\theta}$$

$$= 2\pi \int_0^2 r \sin r^2 dr$$

$$\boxed{r^2 = t \Rightarrow 2rdr = dt}$$

$$= \pi \int_0^4 \sin t dt$$

$$= \pi\left(-\cos t\right)\Big|_0^4 = \pi(1 - \cos 4)$$

Example 2

- Evaluate

$$\iint_R e^{x^2+y^2}\,dxdy$$

where R is the region between the two circles $x^2+y^2=1$ and $x^2+y^2=4$.

- Solution:

$$\iint_R e^{x^2+y^2}\,dxdy = \int_0^{2\pi}\int_1^2 e^{-r^2}\,rdrd\theta$$

$$= \int_0^{2\pi}\left[-\frac{1}{2}e^{-r^2}\right]_1^2 d\theta$$

$$= \pi(e^{-1}-e^{-4})$$

Example 3

- The function exp(-x^2) has no elementary anti-derivative. But we can evaluate $\int_{-\infty}^{\infty} e^{-x^2} dx$ by using the theory of double integrals.

$$\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right) = \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2} dy\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy$$

- Now transform to polar coordinates $x=r\cos\theta$, $y=r\sin\theta$.

$$\left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)} dxdy$$

$$= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2} r dr d\theta$$

$$= \int_{0}^{2\pi}\left[-\frac{1}{2}e^{-r^2}\right]_0^{\infty} d\theta = \pi$$

- Hence

$$\boxed{\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}}$$

3.1.2.3 Generations of Continuous Random Variables

Normal (Gaussian) Distribution

- If $Z \sim N(\mu,\sigma^2)$, its pdf is given by

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(z-\mu)^2}{2\sigma^2} \right\}, \quad -\infty < z < \infty$$

where μ is the mean and σ^2 the variance of the distribution.

- We consider only generation from N(0,1) (standard normal variables), since any random $Z \sim N(\mu,\sigma^2)$ can be represented as $Z=\mu+\sigma X$, where X is from N(0,1).
- Box and Müller Algorithm:
  - Let X and Y be two independent standard normal random variables, so (X,Y) is a random point in the plane.
  - Then the pdf of the two random variable, $f(x,y)$ can be expressed as

$$f(x,y) = \left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right)\left(\frac{1}{\sqrt{2\pi}} e^{-y^2/2} \right)$$

$$= \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

Box and Müller Algorithm (Contd.)

- When $x=r\cos\theta$ and $y=r\sin\theta$, the pdf, $f(r,\theta)$ becomes

$$f(x,y)dxdy = f(r\cos\theta, r\sin\theta)\left|\frac{\partial(r\cos\theta, r\sin\theta)}{\partial(r,\theta)}\right| dr d\theta$$

$$= e^{-r^2/2} r dr d\theta$$

$$\Rightarrow \quad f(r,\theta) = f_R(r) f_\Theta(\theta);$$

$$f_R(r) = e^{-r^2/2} r, \; f_\Theta(\theta) = \frac{1}{2\pi}$$

- Then r can be sampled by

$$F(r) = \int_0^r e^{-r'^2/2} r' dr' = \left[-e^{-r'^2/2}\right]_0^r = 1 - e^{-r^2/2} = \xi'$$

$$\Rightarrow \quad e^{-r^2/2} = \xi_1 \quad \Rightarrow \quad r = \sqrt{-2\ln\xi_1}$$

- Because θ can be generated from the uniform distribution over $[0,2\pi]$, X and Y can be sampled by

$$\boxed{\begin{aligned} x &= \sqrt{-2\ln\xi_1}\cos(2\pi\xi_2), \\ y &= \sqrt{-2\ln\xi_1}\sin(2\pi\xi_2) \end{aligned}}$$

Example #1 – Sampling from Isotropic Distribution

- Method 1

```
phi=2*PI*RNG->GetRN();
sinP=sin(phi); cosP=cos(phi);
```

- Method 2

```
do {
   C1=2.*RNG->GetRN()-1.;
   C2=2.*RNG->GetRN()-1.;
   C3=C1*C1+C2*C2;
}while(C3>1.);
C4=sqrt(C3); sinP=C1/C4; cosP=C2/C4;
```

Proof of Method 2 for the Isotropic Distribution

- Let's consider a uniform distribution in a disc of radius of 1:

$$f(x,y) = \frac{1}{\pi};\ x^2 + y^2 < 1$$

- Then the corresponding pdf for the polar coordinates becomes

$$f(x,y)dxdy = \frac{1}{\pi}\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| drd\theta = \frac{1}{\pi} rdrd\theta$$

$$\Rightarrow \quad f(r,\theta) = f_R(r) f_\Theta(\theta);$$

$$f_R(r) = 2r,\ f_\Theta(\theta) = \frac{1}{2\pi},\ r \in [0,1)$$

- Therefore the uniform pdf in a disc can be regarded as the multiplication of $f_R(r)$ and $f_\Theta(\theta)$ where $f_\Theta(\theta)$ follows the isotropic distribution.
- Then from the sampled X and Y in the xy-coordinates, the R and Θ can be calculated by

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta \end{aligned} \quad \Rightarrow \quad \cos\theta = \frac{X}{\sqrt{X^2+Y^2}},\quad \sin\theta = \frac{Y}{\sqrt{X^2+Y^2}}$$

3.1.2.4 Random Variate Generation from Joint Distribution

Linear Transformation

- Let $\mathbf{x}=(x_1, \ldots, x_n)^T$ be a column vector in R^n and $\mathbf{A}$ an $m \times n$ matrix. The mapping $\mathbf{x} \to \mathbf{z}$, with $\mathbf{z}=\mathbf{A}\mathbf{x}$, is called a linear transformation.
- Now consider a random vector $\mathbf{X}=(X_1, \ldots, X_n)^T$, and let

$$\mathbf{Z} = \mathbf{A}\mathbf{X}$$

  Then $\mathbf{Z}$ is a random vector in R^m.
- Let's see how the expectation vector and covariance matrix of $\mathbf{Z}$ are transformed.

$$\boldsymbol{\mu}_{\mathbf{Z}} \equiv E[\mathbf{Z}] = E[\mathbf{A}\mathbf{X}] = \mathbf{A}E[\mathbf{X}] = \mathbf{A}\boldsymbol{\mu}_{\mathbf{X}}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{Z}} &\equiv E[(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})^T] \\
&= E[(\mathbf{A}\mathbf{X}-\mathbf{A}\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{A}\mathbf{X}-\mathbf{A}\boldsymbol{\mu}_{\mathbf{X}})^T] \\
&= \mathbf{A}E[(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X}-\boldsymbol{\mu}_{\mathbf{X}})^T]\mathbf{A}^T \\
&= \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{A}^T
\end{aligned}$$

Transformation of PDF

- For a linear transformation **z**=**Ax**,

$$f_Z(\mathbf{z}) = \frac{f_X(\mathbf{A}^{-1}\mathbf{z})}{|\mathbf{A}|}$$

- For general transformations

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \rightarrow \begin{pmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_n(\mathbf{x}) \end{pmatrix} \text{ and } \mathbf{z} = \mathbf{g}(\mathbf{x})$$

, the pdf function for **z** becomes

$$f_Z(\mathbf{z}) = f_X(\mathbf{g}^{-1}(\mathbf{z}))\left|J_Z(\mathbf{g}^{-1})\right|; \;\; J_Z(\mathbf{g}^{-1}) = \begin{pmatrix} \frac{\partial g_1^{-1}}{\partial z_1} & \cdots & \frac{\partial g_1^{-1}}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n^{-1}}{\partial z_1} & \cdots & \frac{\partial g_n^{-1}}{\partial z_n} \end{pmatrix} \equiv \frac{\partial(g_1^{-1},\cdots,\partial g_n^{-1})}{\partial(z_1,\cdots,z_n)}$$

Standardization

- Let X~N(0,1). Then X has density f_X given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- Now consider the transformation Z=μ+σX. Then Z has density

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

  In other words, Z~N(μ,σ^2).

- If Z~N(μ,σ^2), then (Z-μ) /σ ~ N(0,1). This procedure is called *standardization*.

Standardization of Joint PDF

- We now generalize this to n dimensions. Let $X_1, \ldots, X_n$ be independent and standard normal random variables. The joint pdf of $\mathbf{X}=(X_1, \ldots, X_n)^T$ is given by

$$f_{\mathbf{X}}(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}$$

- Consider the affine transformation (that is, a linear transformation plus a constant).

$$\mathbf{Z} = \boldsymbol{\mu} + \mathbf{BX}$$

- Then the expectation and covariance becomes

$$\boldsymbol{\mu}_{\mathbf{Z}} \equiv E[\mathbf{Z}] = E[\boldsymbol{\mu} + \mathbf{BX}] = E[\boldsymbol{\mu}] + \mathbf{B}E[\mathbf{X}] = \boldsymbol{\mu}$$

$$\begin{aligned}\boldsymbol{\Sigma}_{\mathbf{Z}} &\equiv E[(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})(\mathbf{Z}-\boldsymbol{\mu}_{\mathbf{Z}})^T] \\ &= E[(\boldsymbol{\mu}+\mathbf{BX}-\boldsymbol{\mu})(\boldsymbol{\mu}+\mathbf{BX}-\boldsymbol{\mu})^T] \\ &= E[(\mathbf{BX})(\mathbf{BX})^T] \\ &= \mathbf{B}E[\mathbf{XX}^T]\mathbf{B}^T \\ &= \mathbf{BIB}^T = \mathbf{BB}^T\end{aligned}$$

- Any random vector of the form of $\mathbf{Z} = \boldsymbol{\mu} + \mathbf{BX}$ is said to have a jointly normal or multivariate normal distribution.

Jointly Normal Random Variables

- Conversely, given a covariance matrix Σ, there exists a unique lower triangular matrix

$$B = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix}$$

such that $\Sigma = \mathbf{BB}^T$. This matrix can be obtained efficiently via the Cholesky decomposition.

Cholesky Square Root Method

- Let Σ be a covariance matrix. Then we wish to find a matrix **B** such that Σ=**BB**^{T}.
- The *Cholesky square root method* computes a lower triangular matrix **B** via a set of recursive equations as follows:

$$\mathbf{Z} = \boldsymbol{\mu} + B\mathbf{X}; \quad B = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix}$$

$$\Rightarrow \quad Z_1 = b_{11}X_1 + \mu_1 \Rightarrow \sigma^2[Z_1] = b_{11}{}^2 \Rightarrow b_{11} = \sqrt{\sigma^2[Z_1]}$$

$$Z_2 = b_{21}X_1 + b_{22}X_2 + \mu_2 \Rightarrow \sigma^2[Z_2] = b_{21}{}^2 + b_{22}{}^2$$

$$\begin{aligned} \mathrm{cov}[Z_1, Z_2] &= E[(Z_1 - \mu_1)(Z_2 - \mu_2)] \\ &= E[((b_{11}X_1 + \mu_1) - \mu_1)((b_{21}X_1 + b_{22}X_2 + \mu_2) - \mu_2)] \\ &= E[b_{11}X_1(b_{21}X_1 + b_{22}X_2)] \\ &= b_{11}b_{21} \end{aligned}$$

$$\Rightarrow b_{21} = \frac{\Sigma_{12}}{b_{11}} = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}}}, \quad b_{22} = \sqrt{\Sigma_{22} - \frac{\Sigma_{21}{}^2}{\Sigma_{11}}}$$

Cholesky Square Root Method (Contd.)

- Generally, the b_{ij} can be found by

$$b_{ij} = \frac{\Sigma_{ij} - \sum_{k=1}^{j-1} b_{ik} b_{jk}}{\sqrt{\Sigma_{jj} - \sum_{k=1}^{j-1} b_{jk}{}^2}}$$

where by convention,

$$\sum_{k=1}^{0} b_{ik} b_{jk} = 0, \ \ 1 \le j \le i \le n$$