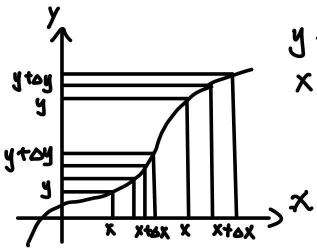
Change of Variable (Transformation)



$y = h(x)$

$x = h^{-1}(y) = g(y)$

$X \sim f_X(x) \qquad Y \sim f_Y(y)$

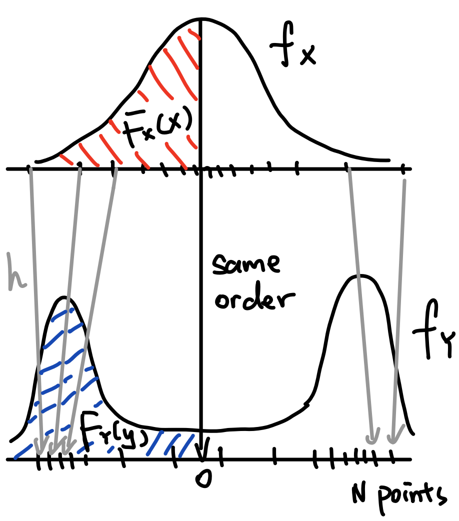
h
g
$y = h(x)$
$x = g(y)$

$$P(\underline{X \in (x, x+\Delta x)}) = P(\underline{Y \in (y, y+\Delta y)})$$

equivalent statements

$$\Rightarrow \boxed{f_X(x)\Delta x = f_Y(y)\Delta y}$$

$$f_Y(y) = f_X(x)\frac{\Delta x}{\Delta y} = f_X(g(y))\,|g'(y)| \quad \left(x = g(y),\ \frac{dx}{dy} = g'(y)\right)$$



order preserving

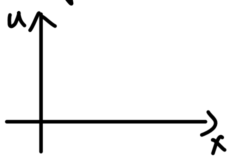
$$P(X \le x) = P(Y \le y)$$

$$\boxed{F_X(x) = F_Y(y)}$$

$$y = F_Y^{-1}(F_X(x)) = h(x)$$

Inversion Method :

Change notation



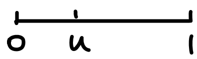
$U \sim \text{Unif}[0,1]$

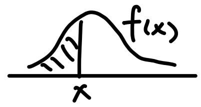
$X \sim f(x)$

$F(x) = P(X \le x)$

$= \int_{-\infty}^{\infty} f(x)\,dx$

$\Rightarrow P(U \le u) = P(X \le x)$





$u \quad = \quad F(x)$

$\Rightarrow \quad x = F^{-1}(u)$

Goal: generate $X \sim f(x)$

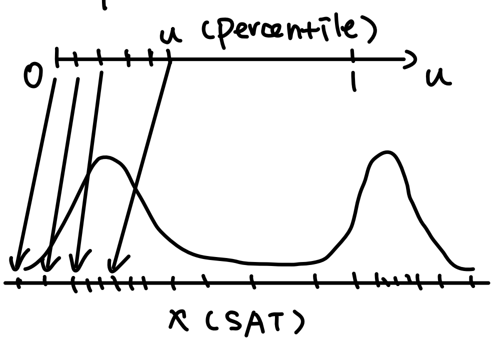
Method:

compute $F(x) = \int_{-\infty}^{x} f(x)\,dx$

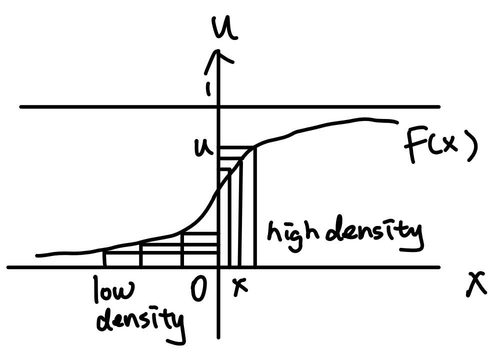
generate $U \sim \text{Unif}[0,1]$

rewrite $X = F^{-1}(U)$

Ex. percentile $\longleftrightarrow$ SAT score



order preserving



density = slope = $F'(x) = f(x)$

Back to Normal Distribution

For a R.V. X

$\mu = E(X) \qquad \sigma^2 = Var(X)$

$\sigma = \sqrt{Var(X)}$

Standardization:

$Z = \dfrac{x-\mu}{\sigma}$

$E(Z) = 0$

$Var(Z) = 1$

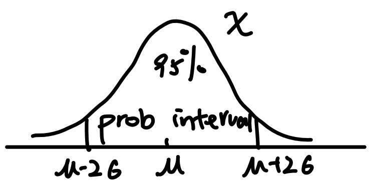
If $X = \varepsilon_1 + \varepsilon_2 + \cdots \varepsilon_n \qquad \varepsilon_i \sim Ber(\frac{1}{2})$

$= \sum_{i=1}^{n} \varepsilon_i \quad$ (ε_i independent)

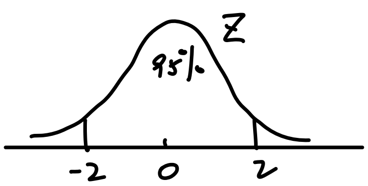
Central Limit Theorem (CLT) (normal approximation)

$Z = \dfrac{x-\mu}{\sigma} \sim N(0,1)$

or $X \sim N(\mu, \sigma^2)$



standardization $\longrightarrow$



Example: $X \sim Bin(n = 100, p = \frac{1}{2})$

$\mu = E(X) = np = 50$

$\sigma^2 = Var(X) = np(1-p) = 25 \qquad \sigma = 5$

95% prob interval: $P(X \in [40, 60]) = 95\%$

Start from binomial distribution :

$$P(X \in [40, 60]) = \sum_{k=40}^{60} P(X=k) = \sum_{k=40}^{60} \frac{\binom{n}{k}}{2^n}$$

$$Z = \frac{x - \frac{n}{2}}{\sqrt{n}/2} = \frac{k - \frac{n}{2}}{\sqrt{n}/2} \quad (X = k \in [40, 60])$$

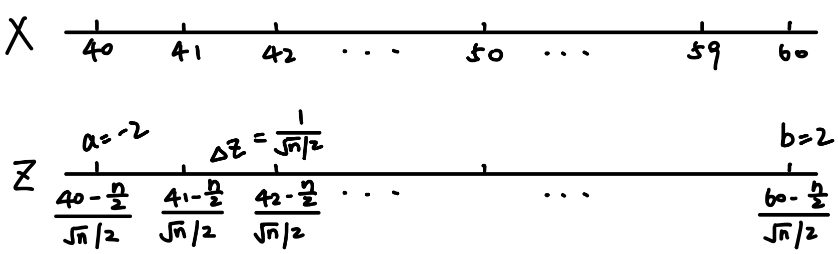
$$\Leftrightarrow k = Z\frac{\sqrt{n}}{2} + \frac{n}{2}$$

$$P(X \in [40, 60]) = \sum_{k=40}^{60} \frac{\binom{n}{k}}{2^n}$$

$$= \sum_{Z \in (a,b)} \frac{\binom{n}{Z\frac{\sqrt{n}}{2} + \frac{n}{2}}}{2^n}$$

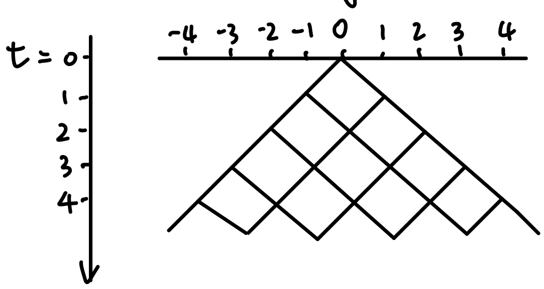
$$= \sum_{Z \in (a,b)} f(Z)\Delta Z \qquad \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}$$

$$= \int_a^b f(Z)\, dZ$$



Random Walk

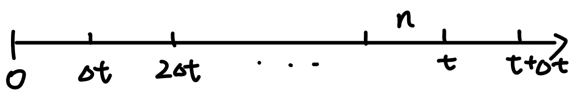
Recall on integer



Similar to Galton Board

Continuous time, Diffusion, Brownian motion

$\Delta t := \frac{t}{n}$

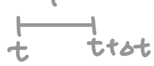


$X(0) \quad X(\Delta t) \quad \cdots \quad X(t) \quad X(t+\Delta t)$

Z_i	-1	$+1$
prob	$\frac{1}{2}$	$\frac{1}{2}$

$$X(t) = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n = \sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} Z_i \Delta X$$

(vs. poisson



$p(\text{head}) = \lambda \Delta t$)

According to Central Limit Theorem:

$$X(t) = \sum_{i=1}^{n} Z_i \Delta X \sim N(0, \sigma^2 t)$$

Q: What is the relationship between x and Δt.

$$E(Z_i) = (-1) \times \frac{1}{2} + (1) \times \frac{1}{2} = 0$$

$$Var(Z_i) = (-1-0)^2 \times \frac{1}{2} + (1-0)^2 \times \frac{1}{2} = 1$$

$$E(X(t)) = E\left(\sum_{i=1}^{n} Z_i \Delta x\right) = \sum_{i=1}^{n} E(Z_i)\Delta x = 0$$

$$Var(X(t)) = \sum_{i=1}^{n} Var(Z_i \cdot \Delta x) = \sum_{i=1}^{n} Var(Z_i)\Delta x^2 = n\Delta x^2 = \sigma^2 t$$

$$\Rightarrow \Delta x^2 = \sigma^2 \frac{t}{n} = \sigma^2 \Delta t$$

$$\Delta x = \sigma \sqrt{\Delta t}$$

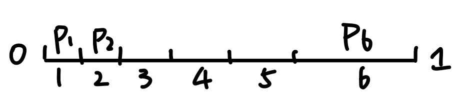
$$\text{Velocity} = \frac{\Delta x}{\Delta t} = \frac{\sigma\sqrt{\Delta t}}{\Delta t} = \frac{\sigma}{\sqrt{\Delta t}} \xrightarrow[0]{\Delta t} \infty$$

Part 2: 2 Random Variables

Notation: $X = (X_1, X_2)$

Discrete: $X \sim p(x)$

biased die



$X = (X_1, X_2)$

$p(x) = p(x_1, x_2)$

$x = (x_1, x_2)$

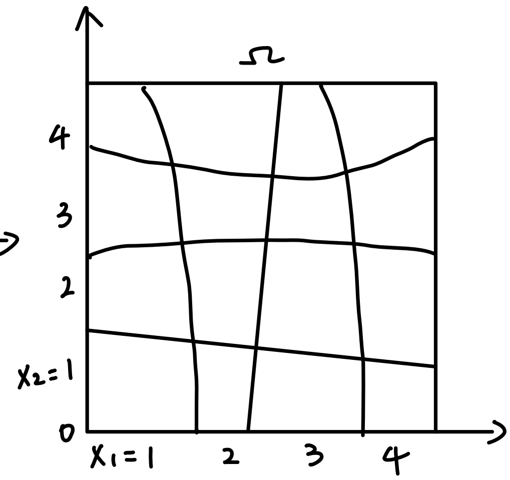
Scalar
↓
vector

Ex.

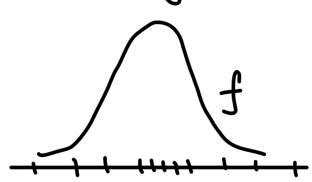
X_1 = eye color \ X_2 = hair color	1	2	3	4
1	p_{11}	p_{12}	p_{13}	p_{14}
2	p_{21}	$\ddots$		$\vdots$
3	p_{31}		$\ddots$	
4	p_{41}	$\cdots$		p_{44}

$\longrightarrow$

$X(w) = (X_1(w) \quad X_2(w))$



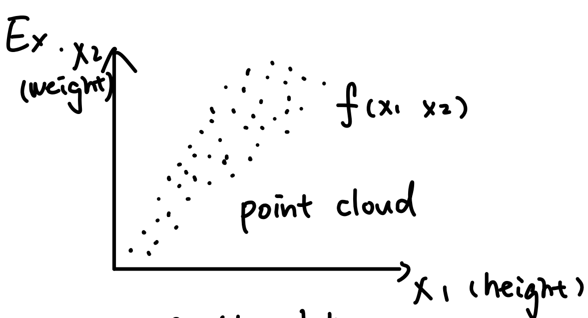
Continuous: $X \sim f(x)$



$X = (X_1 \quad X_2)$

$f(x) = f(x_1 \quad x_2)$

Ex.



scatterplot

3D: $f(x) = f(x_1 \quad x_2 \quad x_3)$