Stochastic process and Gaussian process

--출처: http://enginius.tistory.com/489

1. Stochastic process 란 무엇인가?

Stochastic process 혹은 random process 란 무엇일까? 전자는 수학쪽에서, 후자는 공대쪽에서 주로 사용되는 것 같다. (같은 뜻)

랜덤 프로세스를 정의하기 위해서는 램덤니스, 즉 확률을 먼저 정의해야 한다. 확률이라는 것은 간단하게 생각해서는 주사위를 던졌을 때 각 면이 나올 확률(?)을 의미한다. 쉽게 생각해서 **빈도**라고 하자. 이를 수학적으로 푼 사람이 바로 Andrey Nikolaevich Kolmogorov, 콜모고로프 박사님이시다. 아래의 휜칠하신 분.

이름에서 알 수 있듯이 러시아 수학자이시고, 현대 확률의 개념을 확립하신 분이라고 한다. 확률쪽 논문을 보다보면 이 분 이름이 들어간 이론이 많다. 확률을 정의하기 위해선 **Measure space** 를 먼저 정의해야 한다.

1.1 Measure Space

Measure space 의 정의를 살펴보기 전에, 수학에 대해서 잠깐 얘기해보겠다. 수학, 그중에서도 해석학쪽에서는 low-level 의 수학을 다룬다. 즉 우리가 가지고 있는 특정을 몇 가지 정의하고, 그

정의에서 수많은 내용들을 파생시킨다. 예를 들면 우리가 일반적으로 선형대수를 다루는 공간을 Vector space 라고 한다. 이 공간은 기본적으로 vector sum 과 scalar multiplication 에 닫혀있는 공간으로 정의된다. 이 공간에 거리를 정의할 수 있다. 즉 이 공간에서 임의의 두 점을 뽑았을 때, 둘 사이의 거리를 정의할 수 있으면 해당 공간은 vector space 이다. 이렇듯 수학적으로 공간은, 해당 공간 속에 존재하는 점들의 조건과 그 공간 속에서 정의될 수 있는 연산들로 정의된다.

Measure space 는 (U,B,μ)로 정의된다.

1. U: SET! 모든 가능한 event 나 element 가 모여져 있는 공간을 의미한다.

2. B: Measurable set. U 와 혼동하지 말자. 이는 쉽게 생각해서 U 의 부분집합의 모음이라고 생각하면 된다. 예를 들어서 U={1,2,3}이면, B={∅,{1},{2},{3},{1,2},{1,3},{2,3},{1,2,3}}이다. U 의 원소의 수가 3 개 라면 B 의 원소의 수는 23=8 이다.

3. μ 는 measure 이다. 쉽게 생각해서 U 의 부분집합 혹은 B 의 원소를 실수 R 로 매핑해주는 함수라 생각할 수 있다.

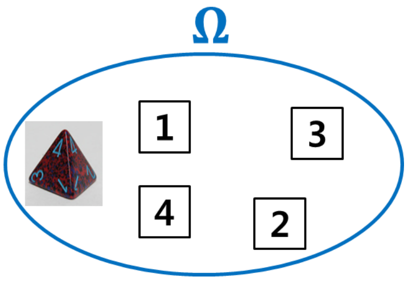
확률이 정의되는 **Probability Space** 는 **Measure Space** 의 특별한 경우이다.

1.2 Probability Space

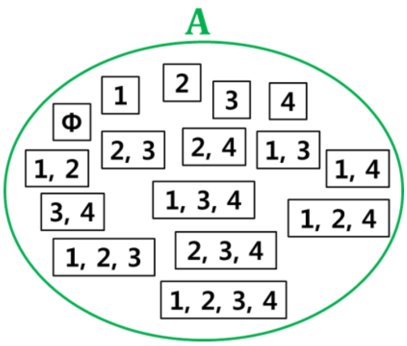
Probability space 는 (Ω,A,P)로 정의된다. 그리고 각 정의는 위와 같고, 다만 P 가 우리가 아는 확률로 0 에서 1 사이 값을 갖는다. Ω 는 sample space 라 한다.

예를 들어서 설명해보겠다. 다음과 같은 1~4 까지 값이 나오는 주사위가 있다고 하자. 이 때 이를 확률로 표현해보면 다음과 같다.

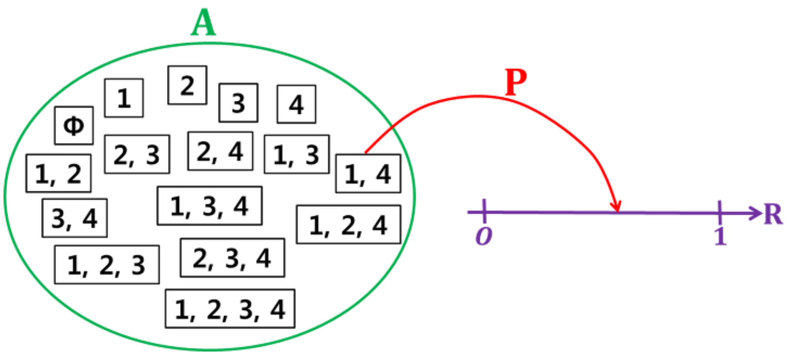
먼저 우리의 Sample space Ω 는 다음과 같다.



그리고 이 공간으로부터 얻을 수 있는 Measurable space 는 다음과 같다.



즉 모든 가능한 경우의 수와 같은 것이다. 이 때 확률은 이 Measurable space 를 [0,1]로 매핑해주는 함수를 의미힌다.



확률을 정의함에 있어서 총 네 가지 axiom 이 쓰인다.

Probability axioms

1. Empty set 의 확률은 0 이다.

2. 확률은 0 보다 크거나 같다.

3. Disjoint sample 들의 union 의 확률은 개별 확률의 합과 같다.

4. Sample space 의 확률은 1 이다.

한가지 주의할 점은 우리가 주사위를 던졌을 때, 1 또는 2 가 나올 확률은 존재하지 않는다. 하지만 P({1,2})는 위의 정의에 따가 P(1)+P(2)로 정의된다. 이렇기 때문에 위에서 Sample space 대신에 Measurable space 에서 확률이 정의된다.

자 여기까지 해서 기본적인 확률에 대한 수학적 정의에 대해서 알아보았다. 이제 random variable 에 대해서 알아보자.

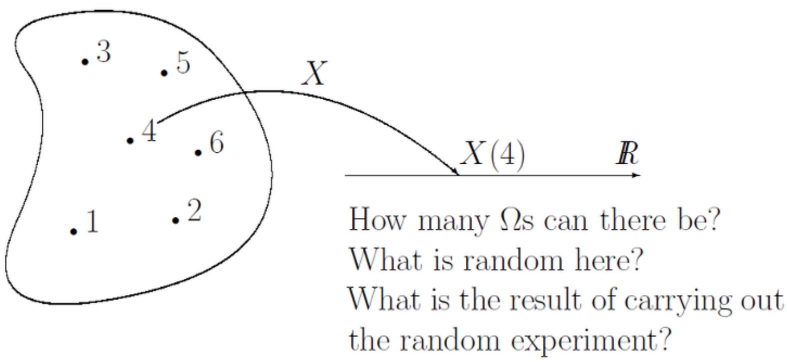
1.3 Random variable

Random variable, 혹은 확률 변수는 Sample space 에서 정의된 어떤 함수이다. 아니 확률 변수가 함수라니 이게 무슨 말인가!

사실 우리가 일반적으로 확률을 의미할 때 이는 다음을 의미한다.

특정 **확률 변수**가 발생할 확률

즉 Sample space 에서 어떤 event 가 발생할 확률을 의미하는 것이 아니 확률 변수가 일어날 확률을 의미한다. 이 두 가지가 어떻게 다를까? 쉽게 생각해서 어떤 임의의 수학적 공간이 있고, 우리는 그 공간에서 우리가 볼 수 있는 공간으로 투영된 값을 읽는 것이다. 다시 말해서 우리가 주사위를 굴려서 얻은 값은 실제 sample space 의 값이 아니라, 이 sample space 에서 어떤 event가 일어나고, 이 event가 확률 변수 X라는 함수를 통해서 우리에게 **보여진** 값이란 의미이다. 이를 **Realization** 이라 한다.



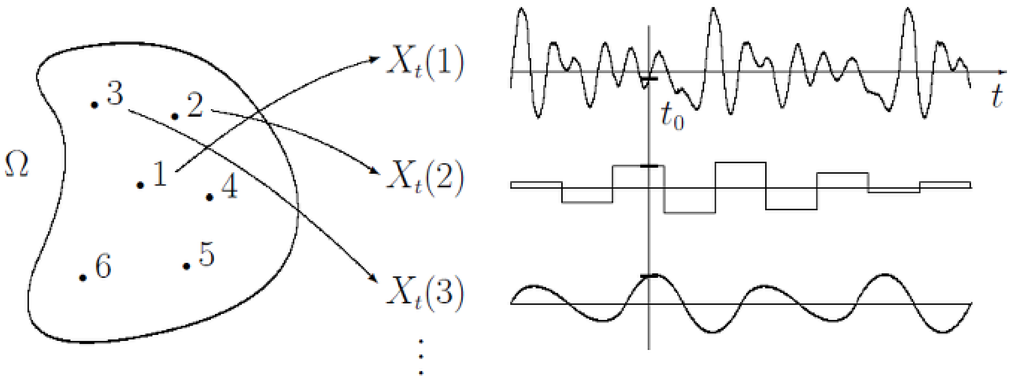
수학적으로 위의 내용을 표현해보면 다음과 같다.

$$P(X\in B)=P(X-1(B))=P(\{w:X(w)\in B\})$$

이 개념이 중요한 이유는 뒤에서 나올 random process 를 정의할 때 혼동이 생기지 않게 하기 위해서이다. 마지막으로 Random process 에 대해서 알아보자.

1.4 Random process

Random process 는 random variable 의 크기를 무한차로 확장시킨 것이라 할 수 있다. 수학적으로 함수는 무한차 벡터를 의미하므로 sample space 의 event 를 어떤 값이 아닌 함수로 realize 한 것을 의미한다.



Random process 는 Xt(w),t∈I 와 같이 쓴다. 이 때 Xt 가 시간 t 에 따라 변하는 random function 을 의미하고, w 는 sample space Ω 에서 뽑은 event 를 의미한다.

1. t 를 고정하면 Xt(w)는 random variable 이 되고,

2. w 를 고정하면 deterministic function 이 된다.

여기서 주의할 점은 t∈I 에서 I 는 임의의 index set 이고, 굳이 1 차원의 시간이 될 필요는 없다. n 차원의 공간으로 확장시킬 수도 있다.

이렇게 정의된 random process 를 표현할 때 많이 사용되는 식이 moment, 즉 mean function 과 covariance function 이다. 물론 RP 가 시간의 개념이 들어가기 때문에 이러한 mean 과 covariance 역시 시간이 들어간다.

Mean function

$$mX(t)=EXt$$

Auto-correlation function (acf)

$$RX(t,s)=EXtXs$$

만약 어떤 RP 가 다음의 특성을 만족한다면 우리는 해당 RP 를 (wide-sense) stationary, wss 라 한다.

1. $mX(t+\tau)=mX(t)=mX$ 즉 mean function 이 상수이다.

2. $RX(t,s)=RX(t-s)$, 즉 acf 과 두 indicating variable 사이의 **거리**의 함수이다.

2. Gaussian process 란 무엇인가?

이번 포스팅의 주 목적인 Gaussian process (GP)에 대해서 생각해보자. GP 는 그 이름에서도 알 수 있듯이 RP 의 일종이다. 그리고, 그 관계가 뭔가 가우시안 적일 것이다. 내가 지금 참고하고 있는 논문 [1]의 노테이션을 사용해서 설명해보겠다.

먼저 RP 는 다음과 같이 정의한다.

Probability space 는 (Ω,F,P)이고, Ω 는 sample space 이다. 확률 변수의 모음으로 $Z(x,w)$로 나타난다. 이 때 x 는 indexing variable 이고, w 는 Sample space 의 event 이다.

이러한 RP 의 mean function 과 covariance function 은 다음과 같이 정의된다.

$$\mu(x)=E(Z(x,w))=\int\Omega Z(x,w)dP(w)$$

$$C(xi,xj)=Cov(Z(xi,w),Z(xj,w))=E((Z(xi,w)-\mu(xi))(Z(xj,w)-\mu(xj)))$$

$$=\int\Omega(Z(xi,w)-\mu(xi))(Z(xj,w)-\mu(xj))dP(w)$$

여기서 부터는 $Z(x,w)$에서 w 의 dependency 를 생략하겠다. 무슨 의민고 하니, 어떤 sample space 로부터 한 event 가 검출되고, 해당 event 로부터 index x 에 의해 어떤 한 값이 결정되는데, 이제부터는 Z 가 index x 에 soley dependent 한다고 보는 것이다. 예를 들어 Z 가 온도를 나타내고, x 가 위치를 나타내면, 온도는 위치에 함수로 보겠다는 의미이다. 그리고 우리는 무한 개가 아닌 다음과 같은 유한 개의 set 을 가정할 것이다.

$$\{Z(x1),Z(x2),\ldots,Z(xn)\}$$

물론 $Z(xi)$는 $Z(x),x\in X$ 의 collection 이다.

Gaussian process 는 유한개의 finite dimensional distribution 이 multivariate normal distribution 을 따르는 RP 를 의미한다. 다시 말해서 유한개의 확률 변수를 뽑아냈을 때 그것들이 모두 joint Gaussian distribution 을 따른다. 그렇기 때문에 GP 는 그것의 mean function 과 covariance function 만 있으면 정의가 된다. 이는 마치 joint Gaussian distribution 이 mean vector 와 covariance matrix 로 정의됨과 마찬가지이다. 또한 covariance function 은 다음의 조건을 만족하는 positive definite function 이어야 한다.

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(x_i,x_j)>0$$

for every n, every collection $\{x_1,x_2,\ldots,x_n\}$ and every vector a.

일반적인 GP 는 mean-zero 를 가정한다. 그리고 covariance function input 들의 차이(separation vector)의 함수이다. 다시 말해서 위에서 정의한 wide-sense stationary 의 조건을 모두 만족한다. 추가로 만약 covariance 가 **distance metric** 의 함수라면 해당 process 는 isotropic 이라고 한다. 이렇게 정의된 covariance function 을 통해서 우리에게 익숙한 kernel (correlation) function 을 유도할 수 있다.

$$R(x_i,x_j)=C(x_i,x_j)\sigma(x_i)\sigma(x_j)$$

$\sigma^2(x_i)=C(x_i,x_i)$ 는 variance function 이다.

wss 를 만족하는 Gaussian process 를 **stationary Gaussian process** 라 한다. 물론 **non-stationary GP** 도 있을 것이다. 이는 우리의 kernel function 이 거리의 함수가 아님을 의미한다.

[1] Paciorek, C., and M. Schervish. "Nonstationary covariance functions for Gaussian process regression." Advances in neural information processing systems 16 (2004): 273-280.