수포자도 도전해 볼 만한

Mathematics in DeepLearning

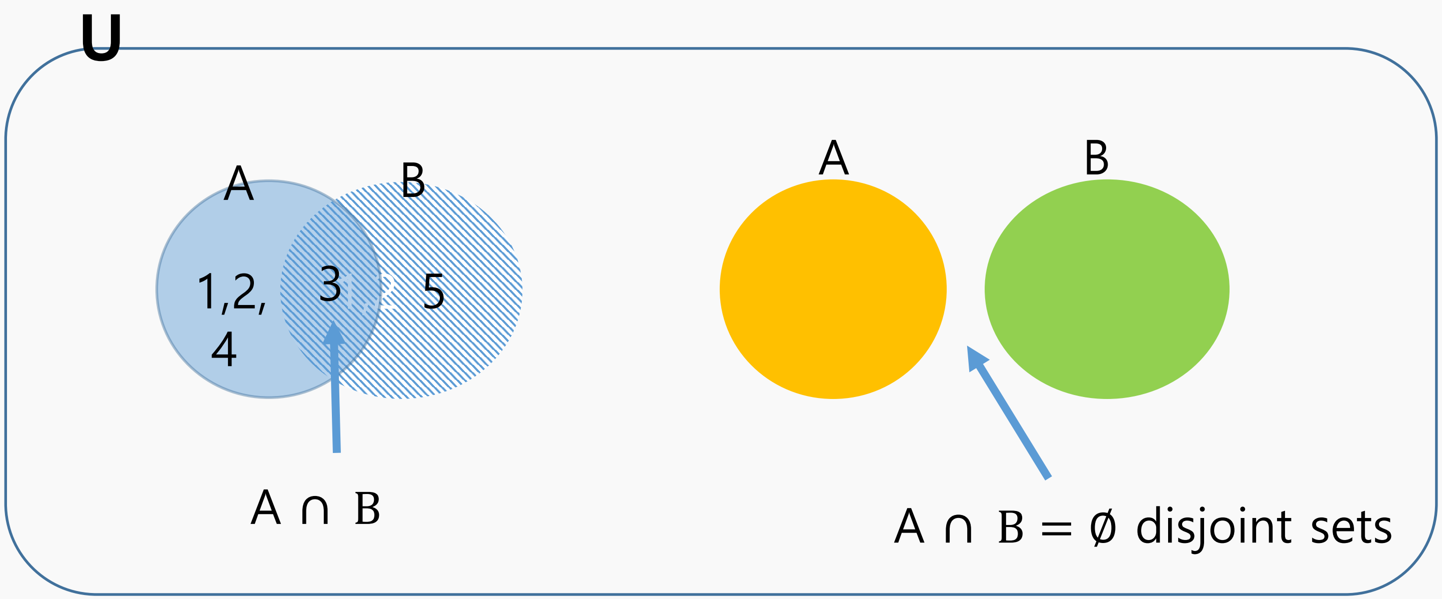
Lecture6. Probability and Statistics

Juhee Lee Ph.D.
Research Professor @ ewha womans university

Last time

Based on Set theory

- Set, universal set, element
- Power set, partition,



Probability

Definition **Sample space (or State space) Ω** is the collection of all possible outcomes under consideration.

Definition An **Event** is a subset $A \subset \Omega$ of the sample space

Probability

Definition A set function $\mathbf{P}$ defined on the set of subsets of Ω is called a probability measure
If it satisfies these 3 conditions for $\mathrm{A}, \mathrm{B} \subset \Omega$

(i) $\mathrm{P}(\mathrm{A}) \geq 0$
(ii) $\mathrm{P}(\emptyset) = 0, P(\Omega) = 1$
(iii) $\mathrm{A} \cap \mathrm{B} = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

set function: assigning a number of a set
Ex. Cardinality, length

Review of Real analysis

<u>Definition</u> A collection $\mathcal{F}$ of subsets of a set Ω is called an **algebra**

(finite union, finite intersection**)**

If

$(i)\ \mathrm{A^c} = \Omega \backslash \mathrm{A} \in \mathcal{F}\ for\ any\ A \in \mathcal{F}$

$(ii)\ A_{1,}A_2 \in \mathcal{F},\ \Rightarrow\ \bigcup_{n=1}^{2} A_n \in \mathcal{F}$

$(iii)\ A_{1,}A_2 \in \mathcal{F},\ \Rightarrow\ \bigcap_{n=1}^{2} A_n \in \mathcal{F}$

Review of Real analysis

Definition A collection $\mathcal{F}$ of subsets of a sample space Ω is called a **σ-algebra**

If

(i) $A^c = \Omega \setminus A \in \mathcal{F}$ *for any* $A \in \mathcal{F}$

(ii) $A_n \in \mathcal{F},\ n = 1,2,,\ldots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

(iii) $A_n \in \mathcal{F},\ n = 1,2,,\ldots \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$

A pair $(\Omega, \mathcal{F})$ is called a **measurable space**

- Algebras (or σ-algebra) are the natural domain of definition of finitely-additive (σ-additive) measure.
- The Lebesgue measurable sets of R form a σ-algebra

$\mathcal{B}(R^n)$: Borel set in R^n

- the smallest σ-algebra generated by the open sets in R^n

- Let T be an arbitrary set. $X = R^T$

$A = \{w \in R^T : (w(t_1), \cdots, w(t_n)) \in E\}$

n is an arbitrary natural number

E an arbitrary Borel set of R^n

$t_1, \cdots, t_n$ an arbitrary collection of distinct elements of T

A an arbitrary subsets in R^T

In the random processes a **probability measure** is defined only on an algebra of this type and extended to the **σ-algebra generated by A**

Probability space

<u>Definition</u> Let $(\Omega, \mathcal{F})$ be a measurable space

A mapping $\boldsymbol{\mu} : \boldsymbol{\mathcal{F}} \to \boldsymbol{R} \cup \{\infty\}$ is called a **measure**

If

(i) $\mu(\mathrm{A}) \geq 0$

(ii) $\mu(\emptyset) = 0, P(\Omega) = 1$

(iii) For any $\mathrm{A}_1, \cdots, A_n$ of mutually disjoint sets in $\mathcal{F}$,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

$(\mu, \Omega, \mathcal{F})$: measure space

Probability is a measure $\mu(\Omega) = 1$, normalized measure

Random Variable

Definition Let $(\mu, \Omega, \mathcal{F})$: measure space.

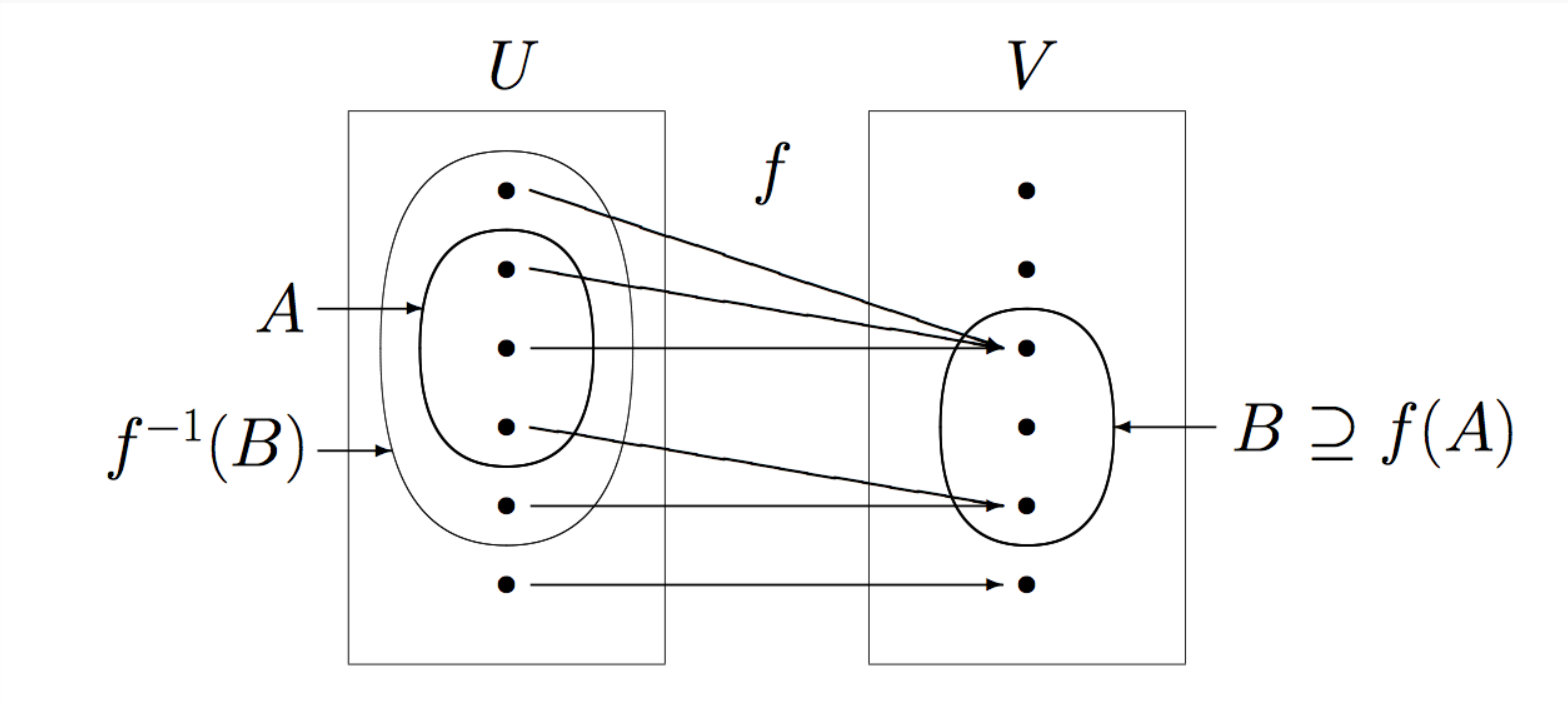
A function $\boldsymbol{X}: \boldsymbol{\Omega} \to \boldsymbol{R^n}$ is said to be **measurable** if $X^{-1}(B) = \{w \in \Omega : \mathrm{X(w)} \in B\} \in \mathcal{F}$

for any $B \in \mathcal{B}(R^n)$

$(\Omega, \mathcal{F}, P)$: probability space

Definition A real measurable function X : $(\boldsymbol{\Omega}, \mathcal{F}) \to (\boldsymbol{R}^n, \mathcal{B}(R^n))$ is called a **random variable**

$$F: U \to V$$



Bayesian Deep Neural Networks (Elementary mathematics by Sungjoon choi, slide)

Random Variable

Discrete random variable :

a set $\{x_1 : i = 1, \ldots\}$ with $\sum P(X = x_i) = 1$

Probability mass function

$P_X(x) = P(X = x)$ *with*

1. $0 \leq P_X(x) \leq 1$
2. $\sum P_X(x) = 1$
3. $P(X \in B) = \sum_{x \in B} P_X(x)$

Random Variable

Continuous random variable :

an integral function $\mathrm{f}_{\mathrm{X}}(\mathrm{x})$ with

$$P(X \in B) = \int_B f_x(x)dx.$$

Probability density function

$$f_x(x) = \lim_{\Delta x \to 0} \frac{P(x < X < x + \Delta x)}{\Delta x}$$

with

1. $f_x(x) \geq 1$
2. $\int f_x(x)dx = 1$
3. $P(X \in B) = \int_B f_x(x)dx.$

$(\Omega, \mathcal{F}, P)$: probability space

Definition The **conditional probability** $P\ (A|B)$ of A given B is defined by

$$P\ (A|B) = \frac{P(A \cap B)}{P(B)} \text{ for } \forall A \in \mathcal{F}$$

If A and B are independent

$$P\ (A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

Bayes' Theorem

$(\Omega, \mathcal{F}, P)$: probability space

$A, B \in \mathcal{F}, P(B) \neq 0$

$$P(A|\,B) = \frac{\mathrm{P(B|A)P(A)}}{\mathrm{P(B)}}$$

Expectation

$(\Omega, \mathcal{F}, P)$: probability space

$\Omega = \{\mathrm{w}_1, \mathrm{w}_2, \dots, w_n\}, P(w_i) = P_i\ , i = 1, \dots, n$
Definition $X: \Omega \to \quad R$ a random vector, $P(\Omega) = \sum_{i=1} P_i = 1$

The expectation of X is defined by
$E(X) = \sum_{i=1}^{n} P_i X(w_i) = \quad \sum_{i=1}^{n} P_i x_i \quad .$
$X(w_i) = x_i$ for $i = 1, \dots, n$

Expectation

Ex. Two coin, X: the number of heads

E(X)=?

- Expectation is a linear functional

$$E(aX + bY) = aE(X) + bE(Y)$$

Expectation

We may assume that $x_1, \cdots, x_n$ for random variable X $such\ that$ $x_1 < \cdots < x_n$

The distribution function F_X is the given by

$F_X(x_i) = P_1 + \cdots + P_{i-1}$ for $x_{i-1} < x < x_i$

with $F_X(x) = \begin{cases} 0, x < x_1 \\ 1, x \geq x_0 \end{cases}$

In particular, $P_i = F_X(x_i) - F_X(x_{i-1}), i = 2, \cdots n,$
$P_1 = F_X(x_1)$

$$E(X) = \sum_{i=1}^{n} P_i x_i = \quad x_1 F_x(x_1) + x_2(F_x(x_2) - F(x_1)) + \cdots$$

Expectation

$$E(X) = \sum_{i=1}^{n} P_i x_i \quad = x_1 F_x(x_1) + x_2(F_x\,(x_2) - F(x_1))$$

$$+ \cdots + x_n(F_x(x_n) - F_x(x_{n-1}))$$

Take, $x_0 < x_1$ and let $F_X\colon [x_0, X_n] \to \quad R$ with

$f_X(x_i) = \quad \frac{F_x(x_n) - F_x(x_{n-1})}{x_i - x_{i-1}}$, $F_x(x_0) = 0$

Expectation

$$E(X)$$

$$= x_1 F_X(x_1) + \sum_{i=2}^{n} x_i \, \frac{F_x(x_n) - F_x(x_{n-1})}{x_i - x_{i-1}} (x_i - x_i)$$

$$= \int x f_x(x) dx$$

Density function

$(\Omega, \mathcal{F}, P)$: probability space , $random\ variable\ X$

Definition If the distribution function F_X is differentiable , then the expectation of X can be written by

$E(X) = \int_{-\infty}^{\infty} x f_x\,(x) dx$ where $f_x\,(x) = \frac{d}{dx} F_X(x)$

We call f_x the **density function** for X

$f_x\,(x)$: non-negatitive, $\int_{-\infty}^{\infty} x f_x\,(x) dx = 1$

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

Consider indep. Rv X and Y $with$ $finitely$ $many$ $values$ $x_1, \cdots, x_n$ and $y_1, \cdots, y_m$ $Then$

$$E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j \, P(X = x_i, Y = y_j) = E(X)E(Y)$$

$$\mathrm{Var}(X) = \mathrm{E}\left(\mathrm{X} - \mathrm{E}(\mathrm{X})\right)^2 = \mathrm{E}(\mathrm{X}^2) - \mathrm{E}(\mathrm{X})^2$$

$$\mathrm{Cov}(X, Y) = \mathrm{E}\left(\mathrm{X} - \mathrm{E}(\mathrm{X})\right)\left(\mathrm{X} - \mathrm{E}(\mathrm{X})\right)$$

$$\mathrm{E}(\mathrm{X}), \mathrm{E}(\mathrm{Y}) < \infty$$

Moments

Definition For a random variable X with density function f_x

$$E(x^n) = \int_{-\infty}^{\infty} x^n f_x\,(x) dx \qquad \text{where } \mathrm{n} \in N$$

Is called the **n-th moment of $\boldsymbol{x}$ about $\boldsymbol{0}$**,

$\mathrm{n} = 3$: skewness

$\mathrm{n} = 4$: kurtosis

Moment generating function

Definition Let X be a random variable.

Suppose that $\exp(\mathrm{t}X) = e^{tX}$ has finite mean for every t in an open interval I with $0 \in I$.

We then define

$$\psi(t) = E(e^{tX}), \mathrm{t} \in I$$

and call ψ the moment generating function (m.g.f) of X

$$\psi(0) = \mathrm{E}(1) = 1$$

Moment generating function

For $\psi(t) = E(e^{tX}), \quad \mathrm{t} \in I$

$$\psi'(t) = E\left(\frac{d}{dt}e^{tX}\right) = E(Xe^{tX}), \mathrm{t} \in I$$

$$\psi'(0) = \mathrm{E(X)},$$

$$\psi''(t) = \mathrm{E}(X^2 e^{tX})$$

$$\psi''(0) = \mathrm{E(X^2)},$$

$$\mathbf{E}(\boldsymbol{X}^{\boldsymbol{n}}) = \boldsymbol{\psi}^{(\boldsymbol{n})}(\mathbf{0}), \quad \mathbf{n} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \cdots.$$

Moment generating function

<u>Theorem</u> Two distributions are identical

If they have m.g.fs coinciding in an open interval around 0 $(\mathrm{t} \in \mathrm{I})$

Proof) Suppose that X and Y are random variables both taking only possible values in $\{0,1,2,\cdots,n\}$ (range of X,Y).

Further, suppose that X and Y have the same m.g.f for all t.

Moment generating function

$$\sum_{x=0}^{n} e^{tx} P(X = x) = \sum_{y=0}^{n} e^{ty} P(Y = y)$$
$$= \sum_{x=0}^{n} e^{tx} P(Y = x)$$

$$\sum_{x=0}^{n} e^{tx} P(X = x) - \sum_{x=0}^{n} e^{tx} P(Y = x) = 0$$

$$\sum_{x=0}^{n} e^{tx} (P(X = x) - P(Y = x)) = 0$$

Moment generating function

let $e^t = s,\quad c_x = P(X = x) - P(Y = x)$

$$\sum_{x=0}^{n} c_x s^x = 0 \Rightarrow \forall x,\quad c_x = 0$$
$$\Rightarrow \forall x,\ P(X = x) = P(Y = x)$$

X, Y have the same distribution ■

Moment generating function

Theorem $X_1, \cdots, X_n$ independent random variables with m.g.f $\psi_1, \cdots, \psi_n$ define in open intervals $I_1, \cdots, I_n$ containging 0.

The m.g.f of $X = a_1 X_1 + \cdots, + a_n X_n$ is

$$\psi_X(t) = \psi_1(a_1 t) \cdots \psi_n(a_n t)$$

$$\mathrm{t} \in a_1^{-1} I_1 \cap \cdots \cap a_n^{-1} I_n$$

Proof) Lecture note.

Theorem X, Y independent random variables

f, g: Borel measurable function

(i.e., $f, g : R \to R$ measurable function)

$\Rightarrow f(X), g(Y)$: independent

Exponential distribution

<u>Definition</u> A random variable $X: \Omega \to R$ is said to be **exponentially distributed** with parameter $\lambda > 0$ if it has a density function f_X given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, \quad\quad x < 0 \end{cases}$$

i.e., $\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^{\infty} = 1$

$$P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}$$

$$
\begin{aligned}
\psi(t) = E(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \\
&= \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} dx\,, \\
&= \lambda \int_{0}^{\infty} e^{(t-\lambda)x} dx\,, \\
&= \frac{\lambda}{t-\lambda}\left[e^{(t-\lambda)x}\right]_0^{\infty} \\
&= \frac{\lambda}{\lambda - t} \qquad\qquad \text{if } -\infty < t < \lambda
\end{aligned}
$$

$$\psi'(t) = \frac{\lambda}{(\lambda - t)^2},\ \psi''(t) = \frac{2\lambda}{(\lambda - t)^3},$$

$$\psi^{(n)}(t) = \frac{\lambda n!}{(\lambda - t)^{n+1}}$$

$$E(X) = \psi'(0) = \frac{1}{\lambda},\ E(X^n) = \frac{n!}{\lambda^n},$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$E(X) \overset{\text{by def}}{=} \int_0^\infty \lambda x e^{-\lambda x} dx$$

i.e. 함수에 따른 E(X) 보다 m.g.f로 구하면 좀 더 편하다.

Binomial distribution

Definition A random variable $X: \Omega \to R$ is said to have the **Bernoulli distributed** with parameter p for $0 \leq p \leq 1$

if it only takes the values 0 and 1

and $\mathrm{P}(\mathrm{X} = 1) = p$ and $\mathrm{P}(\mathrm{X} = 0) = 1 - p$

$$\psi(t) = E(e^{tX}) = e^{t}P(X = 1) + e^{0}P(X = 0)$$
$$= pe^{t} + 1 - p$$

The binomial distribution with parameter $(\mathbf{n}, \boldsymbol{p})$ is the distribution of the sum $\mathbf{X} = \mathbf{X_1} + \cdots + \boldsymbol{X_n},$
$\mathrm{X_i}$: Bernoulli distribution with parameter p

for $0 \le p \le 1$

The random variable X takes integer values from 0 to n

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0,1,\cdots,n$$

m.g.f

$$\psi(t) = (pe^t + 1 - p)^n$$

$$\psi'(t) = npe^t(pe^t + 1 - p)^{n-1}$$

$$\psi''^{(t)} = n(n-1)p^2e^{2t}(pe^t + 1 - p)^{n-2} + npe^t(pe^t + 1 - p)^{n-2}$$

$$\psi'^{(0)} = np = E(X),$$

$$\psi''(0) = n(n-1)p^2 + np = E(X^2)$$

$$Var(X) = \psi''(0) - \{\psi'(0)\}^2$$

$$= np - np^2 = np(1-p)$$

Poisson distribution

<u>Definition</u> A random variable X that takes integer values $\mathrm{x} = 0,1,\cdots$ is Poisson distribution With parameter $\lambda > 0$ if

$$\mathrm{P(X = x) = e}^{-\lambda}\frac{\lambda^{x}}{x!}, x = 0,1,2,\cdots$$

Poisson distribution

$$\psi(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}\ P(X = x)$$

$$= \sum_{x=0}^{\infty} e^{tx}\ e^{-\lambda}\ \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty}\ \frac{(\lambda e^t)^x}{x!}$$

$$= e^{-\lambda}\ \exp(\lambda e^t)$$

$$\psi'(t) = e^{-\lambda}\ \exp(\lambda e^t)\ \lambda e^t$$

Poisson distribution

$$\psi'(t) = e^{-\lambda}\ \exp(\lambda e^t)\ \lambda e^t$$
$$\psi''(t) = e^{-\lambda}\exp(\lambda e^t)\lambda^2 e^{2t} + \lambda e^{-\lambda}\exp(\lambda e^t)e^t$$
$$\psi'(0) = \lambda = E(X),\ \ \psi''(0) = \lambda^2 + \lambda = E(X^2)$$
$$Var(X) = \lambda$$

Characterization of the Poisson distribution

$X_1, \cdots, X_n$: independent random variable

X_i: Poisson distribution with λ_i

$X = \sum X_i$ Poisson distribution with $\lambda = \sum \lambda_i$

Normal (Gaussian) distribution

<u>Definition</u> A random variable $X: \Omega \rightarrow R$ is said to be **normally distributed** with parameter (m, σ) if it has a density function $\mathrm{f_x}$ given by

$$\mathrm{f_X(x)} = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-m)^2}{2\sigma^2})$$

where $\sigma > 0$ and $\mathrm{m} \in R$

Quiz. Is $\mathrm{f_X(x)}$ a pdf?

Check. $\int_{-\infty}^{\infty} f_X(x)dx = 1,$

Normal (Gaussian) distribution

Definition A random variable $X: \Omega \to R$ is said to be **normally distributed** with parameter (m, σ)

if it has a density function $\mathrm{f_x}$ given by

$$\mathrm{f_X(x)} = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-m)^2}{2\sigma^2})$$

where $\sigma > 0$ and $\mathrm{m} \in R$

$$(\mathrm{m}, \sigma) = (0,1) \Rightarrow \psi(t) = \exp(\frac{t^2}{2})$$

$\psi'(0) = m = E(X)$,
$\psi''(0) - \psi'(0)^2 = \sigma^2 = Var(X)$,

Note.

$X_1, \cdots, X_n$: independent random variable
$\boldsymbol{X_i}$ **is normally distributed** with parameter $(\mathrm{m_i}, \sigma_i)$, $i = 1, \cdots, n$

put $X = a_1 X_1 + \cdots + a_n X_n$

The random variable **X is normally distributed** with (m, σ),

where $\mathrm{m} = \sum a_i \mathrm{m_i}$, $\sigma^2 = \sum a_i^2 \sigma_i^2$

<u>Definition</u> $X_1, \cdots, X_n$: independent random variable with $\mathrm{N}(m, \sigma)$

$$\overline{X_n} := \frac{X_1 + \cdots + X_n}{n} \approx E(X)$$

The **sample mean** $\overline{\boldsymbol{X_n}}$ **is normally distributed** with $(m, \frac{\sigma}{\sqrt{\mathrm{n}}})$,

$* \ X \sim B(n, p)$, n is large $\Rightarrow$ $X \approx N(np, npq)$,

Note.

Boole's Inequality
for $\mathrm{A_i} \subset \Omega$

$$\mathrm{P}(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

Jensen's Inequality
for f: convex function, $E(X) < \infty$, $E(f(X)) < \infty$

$$E(f(X)) \geq f(E(X))$$

Note. $\mu = E(X_1) < \infty, \sigma^2 = Var(X_1) < \infty$

Markov's Inequality
$X \geq 0, \forall a > 0$

$$\mathrm{P}(X \geq a) \leq \frac{E(X)}{a}$$

Chebyshev's Inequality

$$\mathrm{P}(|X - E(x)| \geq a) = \mathrm{P}(|X - E(x)|^2 \geq a^2) \leq \frac{Var(X)}{a^2}$$

Law of large numbers

Suppose $X_1, \cdots, X_n$: independent random variable with same distribution, $\mu = E(X_1) < \infty$

For $n \to \infty$

(weak)

$$\forall \epsilon > 0,$$

$$\mathrm{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| < \epsilon \right) \to 1$$

(Strong)

$$\mathrm{P}\left(\frac{X_1 + \cdots + X_n}{n} \to \mu \right) = 1$$

Law of large numbers

큰 수의 법칙 또는 대수의 법칙, 라플라스의 정리

- 큰 모집단에서 무작위로 뽑은 표본의 평균이 전체 모집단의 평균과 가까울 가능성이 높다.
- 독립적인 시행 횟수 n 이 한없이 증가할때, 표본평균은 $E(X)$ 로 수렴하며 사건 A 가 발생할 빈도는 $P(A)$로 수렴한다.

Convergence in distribution

확률변수 $X_1, X_2, \cdots,$ 와 각각의 확률분포함수 $F_1, F_2, \cdots$ 에 대하여 어떤 확률변수 X 와 확률분포함수 F 가 존재하여,

$$\forall x \in R, \qquad \lim_{n \to \infty} F_n(x) = F(x).$$

즉, 확률분포함수의 수렴을 의미하고, 확률변수들이 같은 확률공간에 있을 필요가 없으며, 분포만 고려된다.

Ex. X is normally distributed $\Rightarrow X_n \sim N(0,1)$

Convergence in distribution

CLT (Central limit theorem)

Suppose $X_1, \cdots, X_n$: independent random variable with same distribution, $\mu = E(X_1) < \infty$, $\sigma^2 = Var(X_1) < \infty$

For $n \to \infty$

$$\mathrm{S_n} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \Rightarrow X_n \sim N(0,1)$$

Convergence in Probability

확률변수 $X_1, X_2, \cdots,$ 와 $\forall \epsilon > 0,$

$$\lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0.$$

같은 확률 공간에 있는 확률 변수들의 수렴을 의미한다.

cf. almost convergence (pointwise),
sure convergence

Theorem If a random variable X has a binomial distribution with parameters (n, p), $X \sim B(n, p)$, then for sufficiently large n, the distribution of the variable $Y \coloneqq \frac{x-m}{\sigma} \sim N(0,1)$ where $\mathrm{m} = \mathrm{np}$, $\sigma^2 = npq$.

Stochastic processes

Consider a $(\Omega, \mathcal{F}, P)$: probability space and a time frame T

We only consider the cases where $T = \{0,1,2, \dots\}$ or $T = [0, \infty)$.

<u>Definition</u> A stochastic process X is a map

$$X : T \times \Omega \to R \quad \text{s.t } X_t(w) = X(t, w)$$

is a random variable on Ω for every $t \in T$.

A stochastic process X is **stationary** if random vectors $(X_{t_1}, \cdots, X_{t_n})$ and $(X_{t_{1+h}}, \cdots, X_{t_n+h})$ have identical distributions $\forall t_1 < t_2 < \cdots < t_n$ and h.

X has **stationary increments** if the random variable $X_t - X_s$ and $X_{t+h} - X_{s+h}$ have identical distribution $\forall t > s \geq 0$ and h.

X has **independent increments** if the random variable $X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independed $\forall\ 0 \leq t_1 < t_2 < \cdots < t_n$ and $n \in N$.

The homogeneous Poisson process

Let $T = [0, \infty)$.

A stochastic process $X : T \times \Omega \to \Omega$ is called the **homogenous** poisson process with intensity $X > 0$ if

(i) $X_0 = 0$ (The process begins in 0)

(ii) $X_t - X_s$ and $X_{t+h} - X_{s+h}$ have identical distributions for stationary increment

(iii) $X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independent $0 < t_1 < t_2 < \cdots < t_n$ and $\forall n \geq 1$

The homogeneous Poisson process

(iv) The distribution of X_t is for $t > 0$ given by

$$P(X_t = \mathrm{k}) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0,1,2,\cdots$$

(X_t: Poissson)

Remark.

$$\begin{aligned} X_{t+h} - X_{s+h} &\sim X_t - X_s \\ &\sim X_{t-s} - X_0 \\ &\sim X_{t-s} \end{aligned}$$

The first jumps of the Poisson process

Let Y_1 denote the first time t_1 at which $X_{t_1} \geq 1$

Let Y_2 denote the second time t_1 at which $X_{t_2} \geq 2$

The probability that X has not jumped before time t is given by.

$$P(Y_1 > t) = P(X_t = 0\ ) = e^{-\lambda t}$$

$P(Y_1 \leq t) = 1 - e^{-\lambda t}$: exponential distribution with λ

Let Y_2 denote the time interval between the first and the second jump of the poisson

$Y_1 + Y_2$; the 1st time when the process $X \geq 2$

Quiz. Show that Y_1, Y_2 independent and exponential distributed random variable.

Entropy

To measure the uncertainty associated with random variable

In general, X has a certain number of outcome x_i, $P(X = X_i\ ) = p_i$. A random variable X has discrete values $\{X_1, \cdots, X_n\}$

Definition **Entropy** is a mapping from the space of probability function to the (non-negative) reals givens by $H_s(X) \coloneqq -\sum_{x \in A\_} p(x) \log_s p(x)$

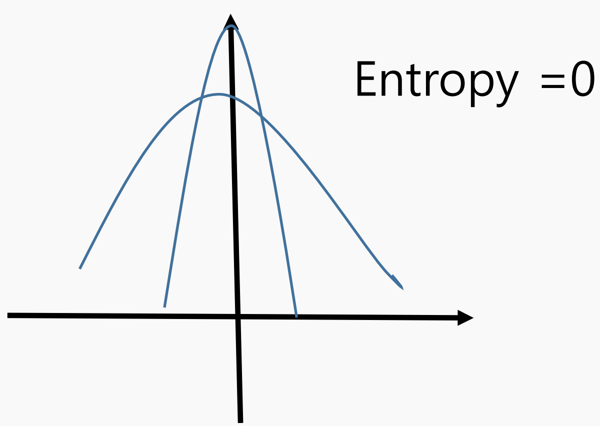
where A is the range of X, $s = 2$ or e

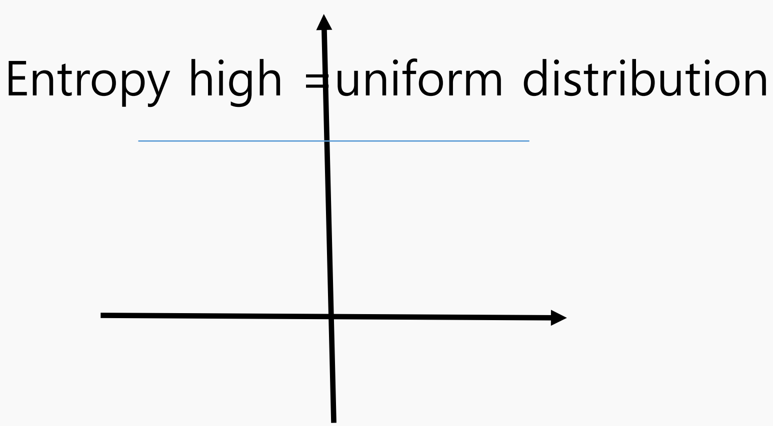
Entropy

Ex. $x \in A$

$$p(X = a) = \begin{cases} 1, & if\ x = a \\ 0, & otherwise \end{cases}$$

Sol) $H_s(X) \coloneqq -log1 = 0$





Q & A