Probability Theory

Hee-il Hahn

Professor
Department of Information and Communications Engineering
Hankuk University of Foreign Studies
hihahn@hufs.ac.kr

Probability distribution

- Marginal probability

  discrete r.v.: $\forall x \in \boldsymbol{x},\ P(\boldsymbol{x} = x) = \sum_y P(\boldsymbol{x} = x, \boldsymbol{y} = y)$

  continuous r.v.: $P(x) = \int P(x, y)dy$

참고:

- random variable: assign a number $\boldsymbol{x}(\zeta)$ to every outcome ζ.

  예) 주사위 던지기에서 $\boldsymbol{x}(i) = 10i$

- $P(\boldsymbol{x} = x) = f_{\boldsymbol{x}}(x)$: probability density function

Probability distribution – cont.

- Conditional probability

$$P(\boldsymbol{y} = y | \boldsymbol{x} = x) = \frac{P(\boldsymbol{x} = x, \boldsymbol{y} = y)}{P(\boldsymbol{x} = x)}$$

- Chain rule of conditional probability

$$P(x^{(1)}, \quad \ldots, \quad x^{(n)}) = P\left(x^{(1)}\right) \prod_{i=2}^{n} P\left(x^{(i)} | x^{(1)}, \quad \ldots, \quad x^{(i-1)}\right)$$

example:

$$P(a, b, c) = P(a|b, c)P(b, c)$$

$$P(b, c) = P(\mathrm{b}|c)P(\,c)$$

$$P(a, b, c) = P(a|b, c)P(\mathrm{b}|c)P(\,c)$$

Probability distribution – cont.

- Independence

$$\forall x \in \boldsymbol{x},\ y \in \boldsymbol{y},\ \ P(\boldsymbol{x} = x, \boldsymbol{y} = y) = P(\boldsymbol{x} = x)P(\boldsymbol{y} = y)$$

참고:

  - uncorrelated : $E[\boldsymbol{xy}] = E[\boldsymbol{x}]E[\boldsymbol{y}]$

$$E[\boldsymbol{x}] = \int xP(\boldsymbol{x} = x)dx, \quad E[\boldsymbol{xy}] = \int xyP(x, y)dxdy$$

  - independent → uncorrelated

- Conditional independence

$$\forall x \in \boldsymbol{x},\ y \in \boldsymbol{y},\ \ \boldsymbol{z} \in z,$$

$$P(\boldsymbol{x} = x, \boldsymbol{y} = y | \boldsymbol{z} = z) = P(\boldsymbol{x} = x | \boldsymbol{z} = z)P(\boldsymbol{y} = y | \boldsymbol{z} = z)$$

Expectation, variance and covariance

- Expectation

$$E_{x\sim P}[f(x)] = \sum_x P(x)f(x)$$

$$E_{x\sim P}[f(x)] = \int P(x)f(x)dx$$

$$E_x[\alpha f(x) + \beta g(x)] = \alpha E_x[f(x)] + \beta E_x[g(x)]$$: linear

- Variance

$$Var\big(f(x)\big) = E[(f(x) - E[f(x)])^2]$$

- Covariance

$$Cov\big(f(x), g(y)\big) = E[(f(x) - E[f(x)])(g(y) - E[g(y)])]$$

Expectation, variance and covariance – cont.

- Covariance

$$Cov\big(f(x), g(y)\big) = E[(f(x) - E[f(x)])(g(y) - E[g(y)])]$$

cf. Correlation

$$Cor\big(f(x), g(y)\big) = \frac{Cov\big(f(x), g(y)\big)}{\sigma_{f(x)}\sigma_{g(y)}}$$

  - *Independence* $\rightarrow$ *Cov = 0*

  - *Covariance matrix*

$$Cov(x)_{i,j} = Cov\big(x_i, x_j\big)$$

$$Cov(x_i, x_i) = Var(x_i)$$

Common probability distributions

- Bernoulli distribution

$$P(\boldsymbol{x} = 1) = \phi$$
$$P(\boldsymbol{x} = 0) = 1 - \phi$$
$$P(\boldsymbol{x} = x) = \phi^x (1 - \phi)^{1-x}$$
$$E[\boldsymbol{x}] = \phi \qquad (E[\boldsymbol{x}] = 1 \cdot \phi + 0 \cdot 1 - \phi\ )$$
$$Var(\boldsymbol{x}) = \phi(1 - \phi)$$
$$Var(\boldsymbol{x}) = E[(\boldsymbol{x} - E[\boldsymbol{x}])^2] = E[\boldsymbol{x}^2] - (E[\boldsymbol{x}])^2 = \phi - \phi^2$$

- Multinoulli distribution
  - Parametrized by a vector $\boldsymbol{p} \in [0, 1]^{k-1}$

    where p_i : prob. of i^{th} state and $p_k = 1 - \mathbf{1}^T \boldsymbol{p}$

Common probability distributions – cont.

- Gaussian distribution

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

  - Central limit theorem : sum of many independent r.v. is approximately normally distributed.

  - Multivariate normal distribution

$$\aleph(x;\ \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sqrt{\frac{1}{(2\pi)^n det(\boldsymbol{\Sigma})}}\, exp\left(-\frac{1}{2}(x-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu})\right),$$

$\boldsymbol{\Sigma}$: covariance matrix

Common probability distributions – cont.

- Exponentional distribution

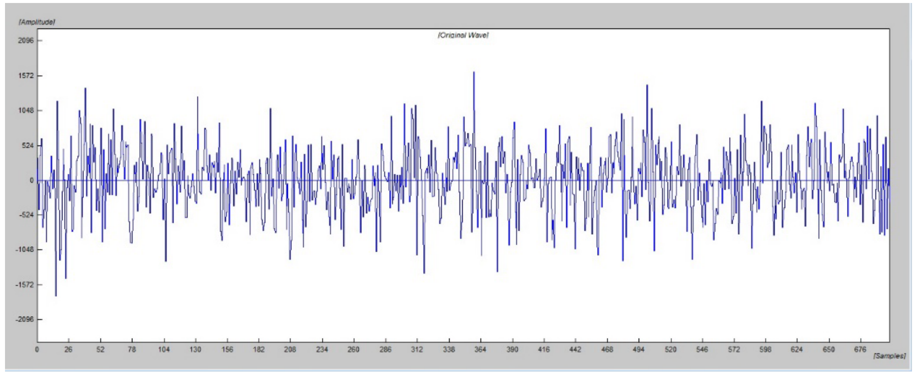
$$p(x;\ \lambda) = \lambda \mathbf{1}_{x \geq 0} exp(-\lambda x)$$

- Laplace distribution

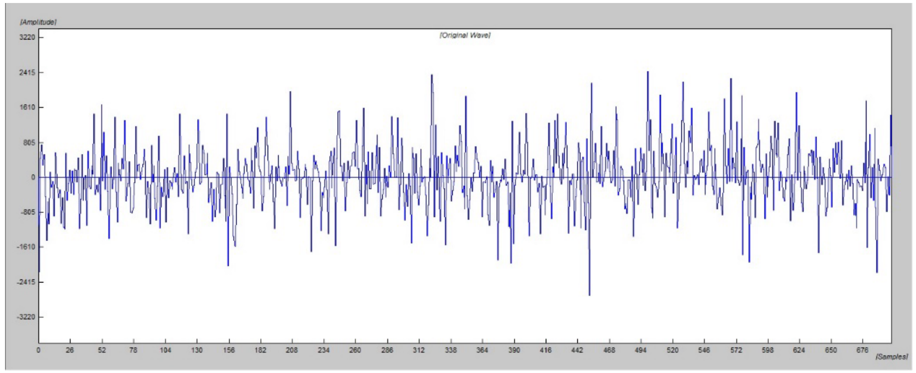
$$Laplace(x;\ \mu,\ \ \gamma) = \frac{1}{2\gamma} exp\left(-\frac{|x-\mu|}{\gamma}\right)$$

Common probability distributions – cont.

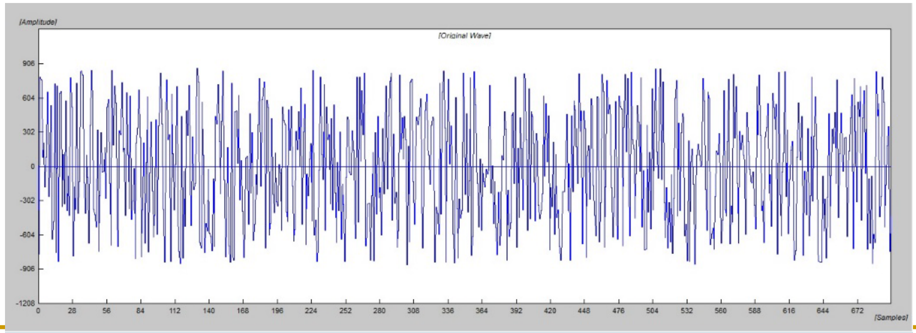
- Gaussian Random Variables



- Laplacian Random Variables



- Uniform Random Variables



Common probability distributions – cont.

- Dirac distribution

$$p(x) = \delta(x - \mu)$$

- Empirical distribution

$$\hat{p}(x) = \frac{1}{m}\sum_{i=1}^{m} \delta\left(x - x^{(i)}\right)$$

Common probability distributions – cont.

- Mixtures of distributions

$$P(x) = \sum_i P(c = i)\, P(x|c = i)$$

$P(c = i)$: multinoulli distribution

- Gaussian mixture model

  $P(x|c = i)$: Gaussian

  $P(c = i)$: prior (a priori) probability

  $P(c \mid x)$: posterior probability

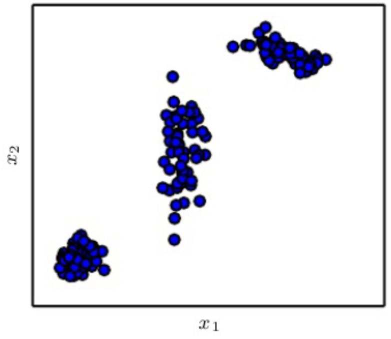


Figure 3.2: Samples from a Gaussian mixture model. In this example, there are three components. From left to right, the first component has an isotropic covariance matrix, meaning it has the same amount of variance in each direction. The second has a diagonal covariance matrix, meaning it can control the variance separately along each axis-aligned direction. This example has more variance along the x_2 axis than along the x_1 axis. The third component has a full-rank covariance matrix, enabling it to control the variance separately along an arbitrary basis of directions.

Useful properties of common functions

- Logistic sigmoid

$$\sigma(x) = \frac{1}{1 + exp(-x)}$$

- Softplus function

$$\varsigma(x) = log\big(1 + exp(x)\big)$$

  - Smoothed version of $x^+ = max(0, x)$

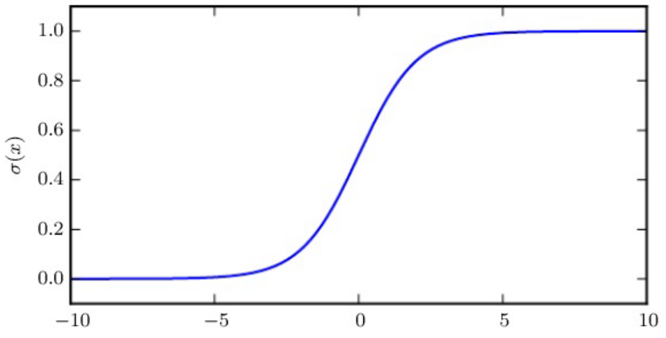


Figure 3.3: The logistic sigmoid function.

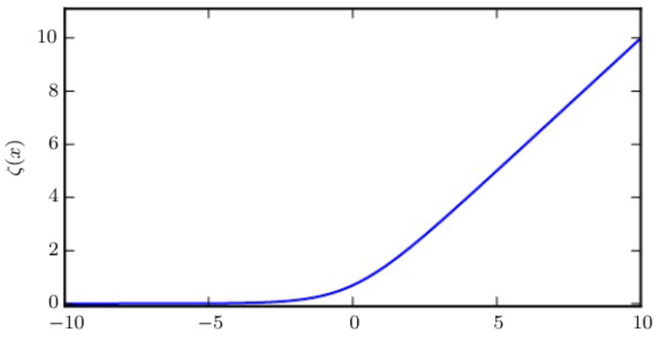


Figure 3.4: The softplus function.

Useful properties of common functions – cont.

$$\sigma(x) = \frac{\exp(x)}{\exp(x) + \exp(0)} \tag{3.33}$$

$$\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x)) \tag{3.34}$$

$$1 - \sigma(x) = \sigma(-x) \tag{3.35}$$

$$\log \sigma(x) = -\zeta(-x) \tag{3.36}$$

$$\frac{d}{dx}\zeta(x) = \sigma(x) \tag{3.37}$$

$$\forall x \in (0, 1),\ \sigma^{-1}(x) = \log\left(\frac{x}{1-x}\right) \tag{3.38}$$

$$\forall x > 0,\ \zeta^{-1}(x) = \log\left(\exp(x) - 1\right) \tag{3.39}$$

$$\zeta(x) = \int_{-\infty}^{x} \sigma(y)dy \tag{3.40}$$

$$\zeta(x) - \zeta(-x) = x \tag{3.41}$$

Bayes' rule

- We need to know $P(x \mid y)$ from $P(y \mid x)$

$$P(x \mid y) = \frac{P(x)P(y \mid x)}{P(y)}$$

$$P(y) = \sum_{x} P(y \mid x)P(x)$$