Department of Systems Engineering

George Mason University

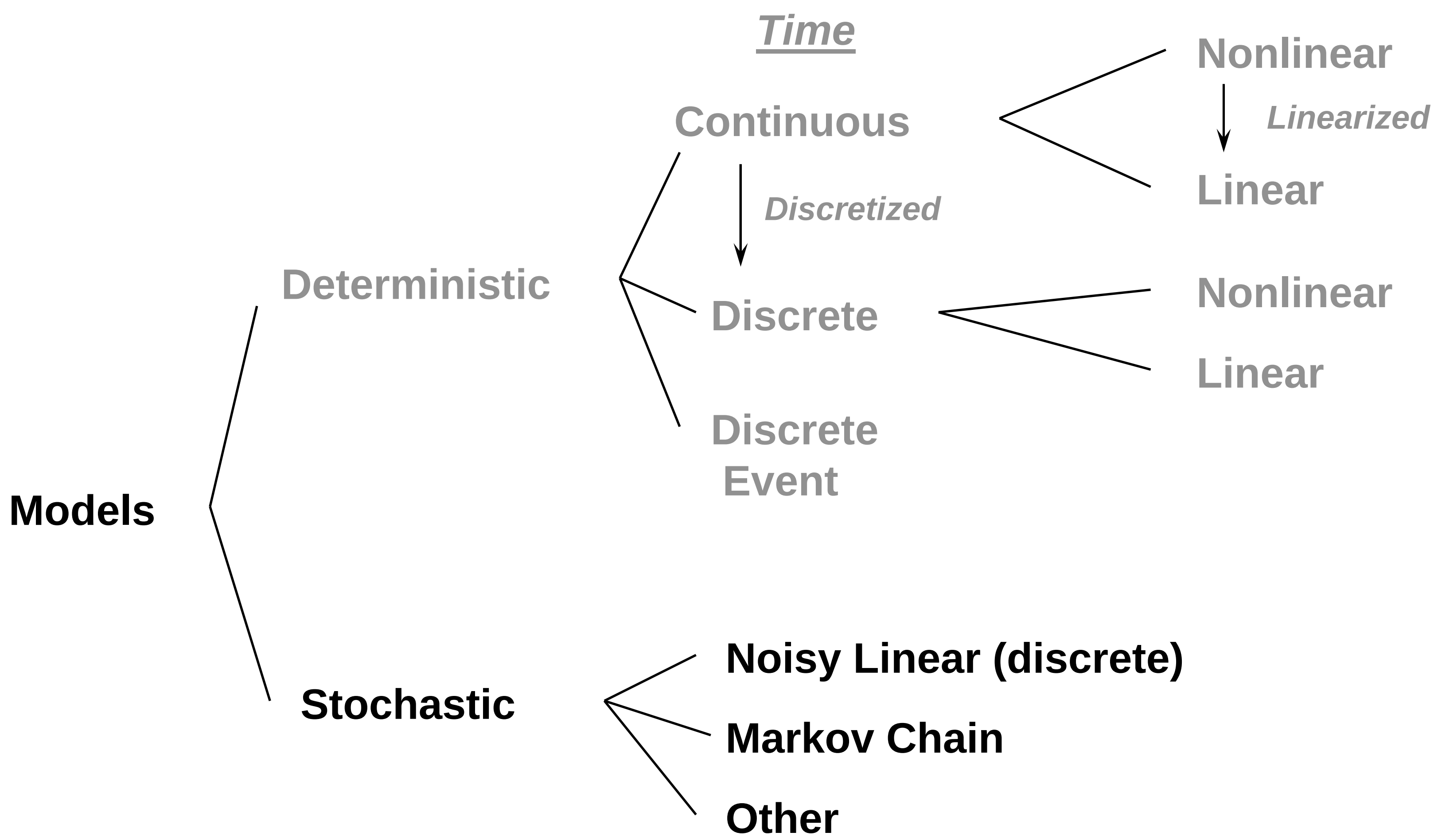
SYST611: Systems Methodology and Modeling #7

Kuo-Chu Chang
Fairfax, Virginia

SYST611: Outline

- Overview of Non-deterministic Systems
- Review of Probability Theory
- Noisy Linear Systems
- Linear Estimation

Taxonomy of Models



Non-Deterministic Systems

- **Model Uncertainties**
  - <u>Deterministic models:</u> knowledge of the state at time *t0* and the input for (*t0,t1*) yields the state and the output at time *t1*
  - <u>Stochastic models:</u> knowledge of the state at time *t0* and the input for (*t0,t1*) yields a probabilistic description of the state and the output at time *t1*
- **Examples**
  - <u>Communication systems:</u> unreliable channel
  - <u>Sensor systems:</u> signal plus noise
  - <u>Resource sharing systems:</u> multi-user queue
- **Mathematical Model**
  - Probability theory and stochastic processes
  - Noisy dynamical systems

Review of Probability Theory

- **Random Variable**
  - A variable x, about which we are uncertain, and to which we can assign a real number
  - <u>Random Vector</u>: A vector of random variables
- **Probability Density Function (pdf)**
  - Define the probability that a random variable x takes on a value in an interval [a b]

$$P(a \le x \le b) = \int_a^b f_x(x)dx \qquad \text{where} \quad \int_{-\infty}^{\infty} f_x(x)dx = 1$$

Cumulative Probability Distribution Function:

$$F_x(x_0) = P(x \le x_0) = \int_{-\infty}^{x_0} f_x(x)dx \quad \text{Note that } P(x > x_0) = 1 - F_x(x_0)$$

Expectation and Variance

- **Expectation (Mean)** $\bar{x} = E(x) = \int_{-\infty}^{\infty} x f_x(x)dx$
- **Variance** $\sigma_x^2 = E[(x-\bar{x})^2] = \int_{-\infty}^{\infty} (x - E[x])^2 f(x)dx = E[x^2] - E[x]^2$
- **Joint pdf** $P(a \le x \le b, c \le y \le d) = \int_a^b \int_c^d f_{x,y}(x,y)dydx$

Cumulative Joint Probability Distribution Function:

$$F_{x,y}(x_0,y_0) = P(x \le x_0, y \le y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} f_{x,y}(x,y)dydx$$

Correlation : $E[xy]$ x and y are orthogonal if E[xy]=0

Covariance : $COV(x,y) = E[(x - E(x))(y - E(y))]$

$$= E[(x-\bar{x})(y-\bar{y})] = E[xy] - \bar{x}\,\bar{y} = \overline{xy} - \bar{x}\,\bar{y}$$

Correlation Coefficient : $\rho_{xy} = \dfrac{\text{COV}[x,y]}{\sigma_x \sigma_y}$, $|\rho| \le 1$

Independent Variables

Probabilistically Independent Random Variables: knowledge about one random variable does not change our uncertainty about the other

$$f_{x,y,\ldots z}(x, y, \ldots z) = f_x(x)\, f_y(y) \cdots f_z(z)$$

or

$$f_x(x \mid y, \ldots, z) = f_x(x)$$

$$\text{independent} \;\Rightarrow\; \text{uncorrelated } (\rho = 0)$$

$$\because \quad E(xy) = E(x)E(y)$$

$$\text{uncorrelated} \;\not\Rightarrow\; \text{independent (except the Gaussian case)}$$

Conditional and Marginal

- **Conditional pdf**

$$f_{x,y}(x|y) = \frac{f_{x,y}(x,y)}{f_y(y)} \Rightarrow f_{x,y}(x,y) = f_{x,y}(x|y)\, f_y(y)$$

- **Marginal pdf**

$$f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)dy$$

$$\text{independent} \quad \Rightarrow \quad \text{conditional} = \text{marginal}$$

$$i.e., \quad f_{x,y}(x|y) = f_x(x)$$

Total Probability Theorem and Bayes Rule

Total Probability Theorem:

$$p(A) = p(A \cap B) + p(A \cap \neg B) = p(A \mid B)\,p(B) + p(A \mid \neg B)\,p(\neg B)$$

$$p(A) = \sum_i p(A \mid B_i)\,p(B_i), \text{ where } \bigcup_i B_i = S,\; B_i \cap B_j = \varnothing\,(i \neq j)$$

$$f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)\,dy = \int_{-\infty}^{\infty} f_x(x \mid y)\,f_y(y)\,dy$$

Bayes Rule:

$$p(B_i \mid A) = \frac{p(A \cap B_i)}{p(A)} = \frac{p(A \mid B_i)\,p(B_i)}{\sum_i p(A \mid B_i)\,p(B_i)},$$

$$\text{where } \bigcup_i B_i = S,\; B_i \cap B_j = \varnothing\,(i \neq j)$$

$$\text{or } f_{x/y}(x \mid y) = \frac{f_{x,y}(x,y)}{f_y(y)} = \frac{f_{y/x}(y \mid x)\,f_x(x)}{\int f_{x,y}(x,y)\,dx}$$

Multivariate Gaussian pdf

Univariate Gaussian(Normal): $f_x(x) = \frac{e^{-((x-\mu)^2/2\sigma^2)}}{\sqrt{2\pi}\sigma} \equiv N(\mu, \sigma^2), \quad -\infty \le x \le \infty$

$E[x] = \mu, \quad V[x] = \sigma^2,$ where σ is the standard deviation

Bivariate Gaussian

$$f_{x,y}(x,y) = \frac{\exp\{\frac{-1}{2(1-\rho_{xy}^2)}[(\frac{x-\mu_x}{\sigma_x})^2 - 2\rho_{xy}(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2]\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}$$

N - Dimensional Multivariate Gaussian

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\}}{(2\pi)^{n/2}|\Sigma|^{1/2}},$$

where $\mathbf{x} = [x_1, \ldots, x_n]', \boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]', E[x_i] = \mu_i, \quad, i = 1,2,\ldots,n$

and $$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{bmatrix}$$

Generating Random Numbers

- **Uniform [0 1]**
  - **Available in most computer packages**
  - **Can be used to generate random variables with other probability distributions**
- **Examples**

Exponential distribution :

$$f_x(x) = \lambda e^{-\lambda x}, \; x > 0 \Rightarrow F_x(x) = 1 - e^{-\lambda x}, \; x > 0$$

$$\text{let } \; y = \mathrm{U}[0 \;\; 1], \; \mathrm{x} = F_x^{-1}(y) = -\frac{1}{\lambda}\ln(1-y)$$

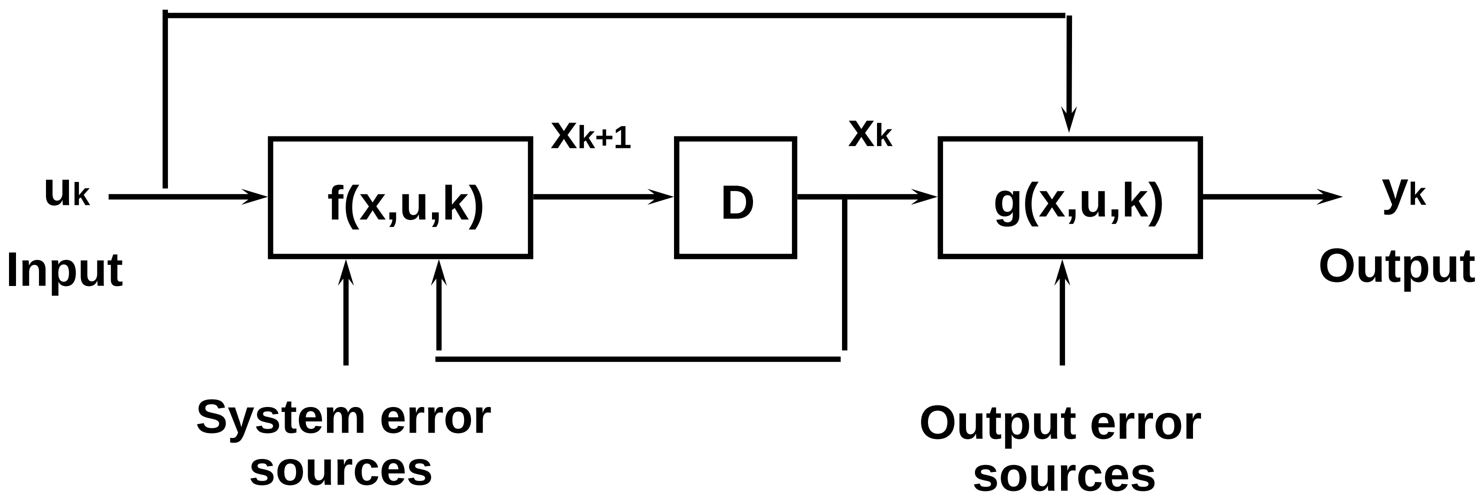
Gaussian distribution :

with two independent uniform random variables $\mathrm{u}_1, \mathrm{u}_2$

then $x = (-2\ln \mathrm{u}_1)^{1/2}\cos(2\pi \mathrm{u}_2)$ and $y = (-2\ln \mathrm{u}_1)^{1/2}\sin(2\pi \mathrm{u}_2)$

are two independent Gaussian variables

Noisy Systems



System error (plant noise): corrupts the relationship of x_k and u_k on x_{k+1}

Output error (observation noise): corrupts the output (observation) y_k

Noisy Discrete Linear Systems

Consider the discrete-time linear dynamic system

$\mathbf{x}(k+1) = F(k)\mathbf{x}(k) + G(k)\mathbf{u}(k) + \mathbf{v}(k) \quad k = 0,1,2,....$

where $\mathbf{x}(k)$ is $\mathrm{n_x}$ – dimensional state vector, $\mathbf{u}(k)$ is $\mathrm{n_u}$ – dimensional **known input** vector, and $\mathbf{v}(k)$ is zero-mean white Gaussian **process noise** (plant noise) with covariance $E[\mathbf{v}(k)\mathbf{v}(k)'] = Q(k)$

The measurement equation is

$\mathbf{y}(k) = \mathrm{H}(k)\mathbf{x}(k) + \mathbf{w}(k) \quad k = 0,1,2,....$

$\mathbf{w}(k)$ is zero-mean white Gaussian **measurement noise** with covariance $E[\mathbf{w}(k)\mathbf{w}(k)'] = R(k)$

$\mathbf{x}(0)$ is assumed to be a Gaussian distributed random variable with known mean and covariance, and independent of $\mathbf{v}(k)$ and $\mathbf{w}(k)$

$\Rightarrow$ **Linear Gaussian** assumption

Discrete System Example

Consider $\mathbf{x}(k+1) = F\mathbf{x}(k) + G\mathbf{u}(k) + \mathbf{v}(k) \quad k = 0,1,2,....$

with $\mathrm{E}[\mathbf{v}(k)] = 0$ and $E[\mathbf{v}(k)\mathbf{v}(k)'] = Q$

By definition, it is a Markov process

$$f_{x_{k+1}}(\mathbf{x}(k+1)|\mathbf{x}(k), \mathbf{x}(k-1), \ldots, \mathbf{x}(0)) = f_{x_{k+1}}(\mathbf{x}(k+1)|\mathbf{x}(k))$$

in fact, $f_{x_{k+1}}(\mathbf{x}(k+1)|\mathbf{x}(k)) = N[F\,\mathbf{x}(k) + G\mathbf{u}(k), Q]$

icbest $\mathbf{x}(k) = F^k\mathbf{x}(0) + \sum_{i=0}^{k-1} F^{k-1-i}G\mathbf{u}(i) + \sum_{i=0}^{k-1} F^{k-1-i}\mathbf{v}(i)$

$$f_{x_{k+1}}(\mathbf{x}(k)) = N[\bar{\mathbf{x}}(k), \Sigma(k)]$$

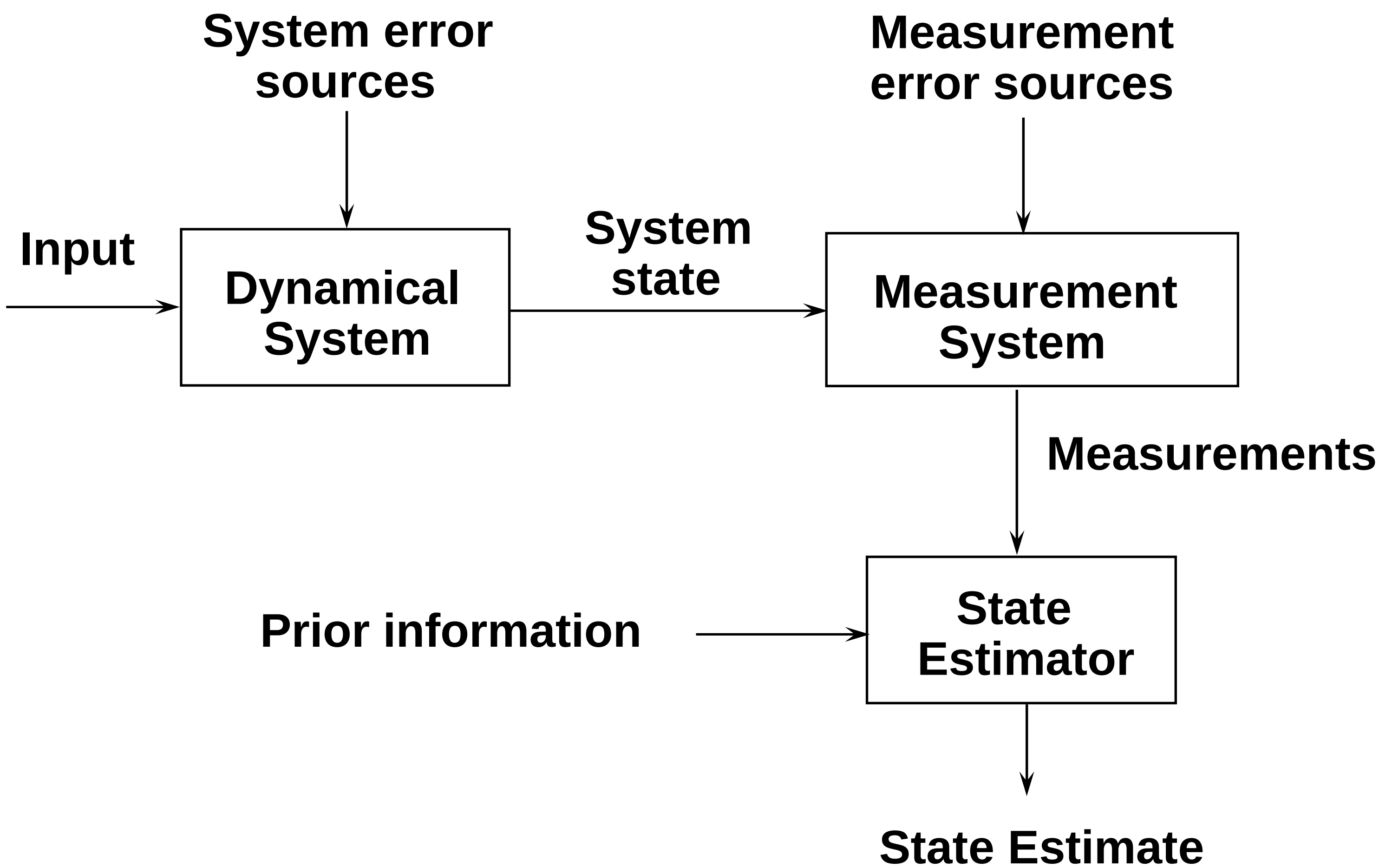
where $\bar{\mathbf{x}}(k) = F^k\mathbf{x}(0) + \sum_{i=0}^{k-1} F^{k-1-i}G\mathbf{u}(i)$

and $\Sigma(k) = \sum_{i=0}^{k-1} F^{k-1-i}Q(F^{k-1-i})'$

Linear Estimation

- **Estimation:** the process of inferring the value of a quantity of interest from indirect, inaccurate, and uncertain observations
  - Determine model parameters for physical system - system identification
  - Determine position and velocity of target - tracking
  - Determine message characteristics from noisy channel - communication
- **Filtering:** the estimation of the state of a dynamic system
  - To obtain "best estimate" from noisy data
  - To "filter out" or to eliminate undesired signal
- **Optimal Estimator:** the algorithm that yields an estimate of interest which minimizes a certain error criterion

State Estimation



Bayesian vs. Non-Bayesian Approach

To estimate parameter x given measurements

$$Z(j) = h[j, x, \omega(j)], \qquad j = 1,2,\ldots.,k$$

with noise $\omega(j)$, find a function (***estimator***)

$$\hat{x}(k) \equiv \hat{x}[k, Z^k], \qquad Z^k = \{Z(j)\}_{j=1}^{k}$$

that estimates the value of x

Bayesian : if prior pdf of x is available

$$\textit{Bayes' Rule}: \quad p(x|Z) = \frac{p(Z|x)p(x)}{p(Z)} = \frac{1}{C} p(Z|x)p(x)$$

Non - Bayesian : if prior pdf of x is unavailable

$$\textit{likelihood function}: \quad \Lambda_Z(x) \equiv p(Z|x)$$

Maximum Likelihood and Maximum A Posterior Estimates

Non – Bayesian :

Maximum Likelihood Estimate (MLE):

$$\hat{x}^{ML}(Z) = \max_x \Lambda_Z(x) = \max_x p(Z|x)$$

Bayesian :

Maximum A Posterior Estimate (MAP):

$$\hat{x}^{MAP}(Z) = \max_x p(x|Z) = \max_x [p(Z|x)p(x)]$$

ML Estimator Example

For example: (parameter estimation)

Given a static parameter x with Gaussian prior

$$p(x) \sim N(x; \bar{x}, \sigma_0^2)$$

with observation model: $z = x + \omega$,

where $\omega \sim N(0, \sigma^2)$, then with ML estimator

$$\Lambda(x) = p(z|x) = N(z; x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-x)^2}{2\sigma^2}}$$

$$\Rightarrow \hat{x}^{ML} = \max_x \Lambda(x) = z$$

MAP Estimator

With MAP estimator, assume x and ω are independent

$$p(x|z) = \frac{p(z|x)P(x)}{p(z)} = \frac{1}{C} e^{-\frac{(z-x)^2}{2\sigma^2} - \frac{(x-\bar{x})^2}{2\sigma_0^2}} \Rightarrow \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\eta(z))^2}{2\sigma_1^2}}$$

where $\eta(z) = \bar{x} + \dfrac{\sigma_0^2}{\sigma_0^2 + \sigma^2}(z - \bar{x}) = \dfrac{\sigma^2}{\sigma_0^2 + \sigma^2}\bar{x} + \dfrac{\sigma_0^2}{\sigma_0^2 + \sigma^2} z$

and $\sigma_1^2 = \dfrac{\sigma_0^2 \sigma^2}{\sigma_0^2 + \sigma^2} = \left(\dfrac{1}{\sigma_0^2} + \dfrac{1}{\sigma^2} \right)^{-1}$

$$\Rightarrow \hat{x}^{MAP} = \eta(z) = \sigma_1^2 \sigma_0^{-2} \bar{x} + \sigma_1^2 \sigma^{-2} z$$

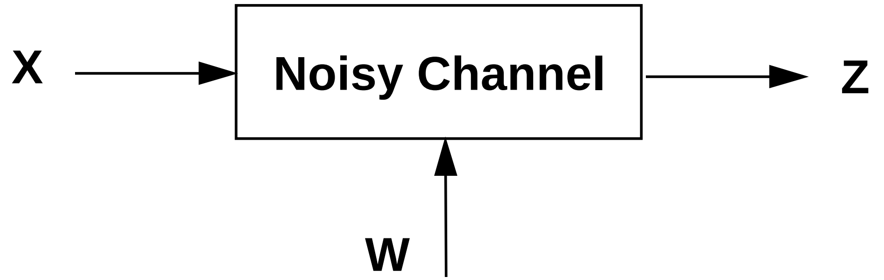
i.e., weighting are proportional to their variance inverse

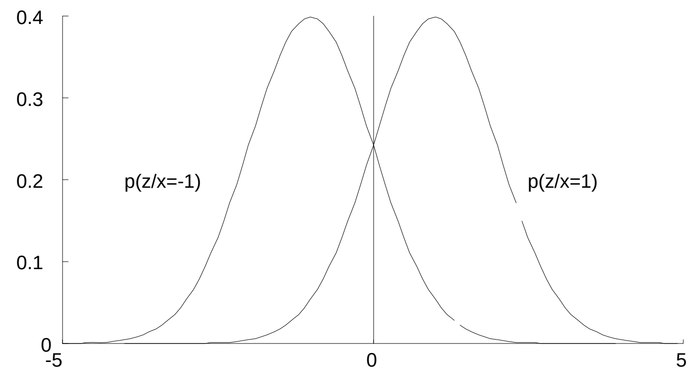
Communication Channel Example

$$z = x + w$$

$$x = \begin{cases} 1 \\ -1 \end{cases}, \quad w = N(0,1)$$

$$\Rightarrow p(z|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}}$$





Maximum likelihood decision policy: if z > 0, x=1 if z < 0, x=-1

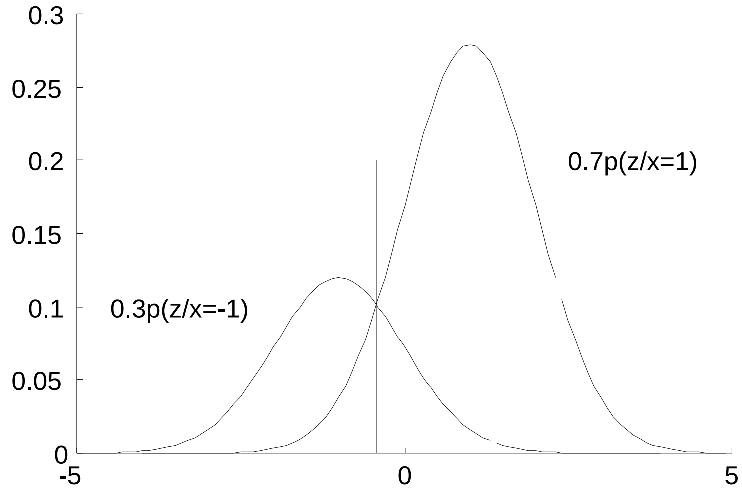
$$\text{Prob(error)} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{\frac{-(t-1)^2}{2}} dt = erf(-1) = 0.1587$$

MAP Estimate Example

$$z = x + w$$

$$P(x=1) = 0.7,\ P(x=-1) = 0.3$$

$$w = N(0,1)$$



$$\text{MAP Estimate} : \max_{x} p(x \mid z) = \max_{x} \frac{1}{C} p(z \mid x) P(x)$$

$$\text{Decision boundary} : p(z \mid x=1) P(x=1) = p(z \mid x=-1) P(x=-1)$$

$$\Rightarrow 0.7 e^{-\frac{(z-1)^2}{2}} = 0.3 e^{-\frac{(z+1)^2}{2}} \Rightarrow z = \ln(3/7)/2 = -0.4236$$

$$\text{Prob(error)} = 0.7 erf(-1.4236) + 0.3[1 - erf(1 - 0.4236)]$$

$$= 0.7 * 0.0773 + 0.3 * [1 - 0.7178] = 0.0541 + 0.0847 = 0.1388$$