3

Bayesian Estimation and Measurement

Consider the robot is about to explore the environment. In order to understand the surrounding environment, first the robot should obtain the sensor information about the environment, for example distances between the robot and the surrounding structures. The sensor information, in this case, is a measurement of the true distance. The problem lies here. In the realm of Robotics, our sensors serve as imperfect observers, providing us with measurements that are, in essence, true essence of the environment involving noises. This “true essence” represents the genuine state of the surrounding environments or events, unblemished by the inherent inaccuracies of the sensors. Our goal is to uncover the genuine nature concealed within the noisy sensor measurements, and achieve the true essence of the surroundings for the robot perception and further robot applications. Simply, our goal is to estimate the true value from the measurements.

In order to discern and achieve this authentic landscape, we turn to the probabilistic estimation theory. Apparently, it is impossible to know the correct noise model of the certain sensors. However this is not the end. We empirically know that the various noise models in nature approximately follow the known distributions such as Gaussian distribution, thereby it is possible to approximate and make sense of the true reality that lies beneath the surface of our sensor measurements. Here Bayesian estimation becomes our guiding light from now.

3.1 Maximum Likelihood Estimation

Suppose the robot is equipped with the time-of-flight (ToF) sensor and is in a stationary state. Consider that the robot is measuring the distance to the structure placed in front. Here, let $x[n]$ be the measurement obtained

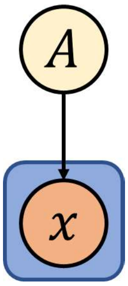

Figure 3.1: Graph of the true value A and the measurement x. The measurement model can be represented as the Graph structure.

from the sensor in time n, and let A be the true distance between the sensor and the structure. Since the measurement of the sensor inevitably involves noise, we cannot say that the measurement $x[n]$ is the true distance A, but we can approximately find the value of A in probabilistic manner.

For this estimation, first we need to assume the noise model (or measurement model) of the sensor with known distributions. In usual, Gaussian distribution plays a main role as the measurement model. Assume that $w[n]$ is the random variable representing the noise occurred in the n^{th} measurement. Let us say the robot obtains N times measurements, then we can express the relation between $x[n]$ and A as the following:

$$x[n] = A + w[n]\,,\;\; n = 0, 1, ..., N-1 \tag{3.1}$$

where

$$w[n] \sim \mathcal{N}\left(0, \sigma^2\right). \tag{3.2}$$

Here σ^2 reflects the uncertainty of the sensor measurement. In this example let us assume that σ^2 is known, which can be obtained empirically in practice. The relation between A and x are depicted as a graph in Fig. 3.1.

Since $w[n]$ is a Gaussian random variable and A is a constant, $x[n]$ is also a Gaussian random variable with the following probability distribution:

$$p(x[n]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x[n]-A)^2}{2\sigma^2}\right], \tag{3.3}$$

which is the *likelihood* of the random variable $x[n]$. In this equation, apparently A is a constant hence not a probabilistic random variable. However, in usual we express A as if a random variable for convenience, because when the value of A is determined, we can say the random variable $x[n]$ follows the above probability distribution which is determined by A. In other words, the distribution of $x[n]$ is conditionally determined by A. In this context, we can also rewrite this likelihood as if the conditional probability:

$$p(x[n]; A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x[n]-A)^2}{2\sigma^2}\right]. \tag{3.4}$$

Apparently, in order to represent the conditional probability the left-hand side of the above equation should be $p(x[n]|A)$. However, we need to note that the likelihood is not a conditional probability in usual because A in our case is a constant, not a random variable. Therefore we simply use a semicolon to express the relation between $x[n]$ and A in the likelihood.

The likelihood involves the noise model of the measurement, thus we can utilize the likelihood as the measurement model (or observation model). Since the robot measures the distance N times, under the assumption that all the measurements are independent to each other we can represent the likelihood of N measurements as the following:

$$\begin{aligned} p(\boldsymbol{x}; A) &= p(x[0], ..., x[N-1]; A) \\ &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \end{aligned} \tag{3.5}$$

where $\boldsymbol{x} = \{x[0], ..., x[N-1]\}$. To find the optimal parameter of A, the first approach we can consider is Maximum Likelihood Estimation (MLE). Suppose we have optimal parameter $\hat{A}$ for the true distance. Since we assume that the measurements $\boldsymbol{x}$ is sampled from the likelihood, this likelihood has the maximum probability with the optimal parameter $\hat{A}$ thus we have the following expression:

$$\hat{A} = \underset{A}{\operatorname{argmax}}\, p(\boldsymbol{x}; A) \tag{3.6}$$

Therefore, the optimal parameter $\hat{A}$ can be calculated by taking the derivation of the likelihood. For convenience, in usual we use log-likelihood for Gaussian distribution, since Gaussian involves quadratic form which is convex and the logarithm is monotonically increasing function so that the optimal point remains the same while in logarithm form. Therefore we have:

$$\hat{A} = \underset{A}{\operatorname{argmax}} \log p(\boldsymbol{x}; A) \tag{3.7}$$

where

$$\begin{aligned}\log p\left(\boldsymbol{x}; A\right) &= -\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x\left[n\right] - A\right)^2 \\ &= const - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x\left[n\right] - A\right)^2 . \end{aligned} \tag{3.8}$$

Note that in the case of Gaussian, we have quadratic form inside of the exponential form when taking logarithm of Gaussian. This will play an important role in further. By taking the partial derivative of the log-likelihood, we have the following:

$$\frac{\partial \log p\left(\boldsymbol{x}; A\right)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x\left[n\right] - A\right). \tag{3.9}$$

Set to zero we finally have:

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x\left[n\right]. \tag{3.10}$$

Therefore, for the measurements of the constant quantities such as the distance between the static robot and the surrounding structure, we can conclude that the optimal parameters can be determined by taking average of the multiple measurement. Note that the noise model (or observation model) here is assumed to be Gaussian.

3.2 Bayesian Mean Square Error

Estimating the true value of the target quantities always poses a great challenge, due to the measurement noise or model uncertainty. In the previous section we handled this uncertainty by assuming the measurement model as Gaussian and by applying MLE. In this section we handle this noise by simply minimizing the error and compare the result to the MLE case. Assume that the robot at a standstill measures the distance to the static object and obtain the measurement $x\left[n\right]$ in time n, and let A be the true distance. Due to the noises from the measurement model, the measurement $x\left[n\right]$ and the true distance A always pose a difference, which is the error that we need to minimize. To compute the error, we define Least square error criterion J as the cost function as the following:

$$J\left(A\right) = \sum_{n=0}^{N-1}\left(x\left[n\right] - A\right)^2 . \tag{3.11}$$

Apparently, the optimal solution $\hat{A}$ of this quadratic form can be given as:

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x\left[n\right], \tag{3.12}$$

which is the same result of the optimal solution from MLE.

Although the least squares method enables us to attain the same solution in Gaussian scenarios, it falls short in providing insights beyond measurement models expressed solely in terms of likelihood.

In robotics, there are instances where various information can be anticipated in advance through pre-existing knowledge of the system model. For example, having prior information about the surrounding environment allows for a rough anticipation of the positions of observed objects. Additionally, knowing the control input provided to the robot enables the prediction of the robot's anticipated location before movement. From this, predictions can be made regarding how far an observed object will be from the robot compared to its position before movement.

To illustrate, consider having prior knowledge about the approximate distance of an object placed in front of a robot. Let the true distance from the robot to the object be denoted as A, and the observed distance from the measurement be denoted as x. Since likelihood only represents the measurement model, given A, it mimics the probability density function (PDF) for the observed value x. Therefore, there is no avenue to incorporate our existing knowledge about the object's position. Here, A is simply a constant parameter.

However, with prior knowledge about the approximate range of distances to the object, we now possess probabilistic information about A, treating it as a random variable. For instance, if we have prior knowledge that A lies between a and b, the PDF for A can be expressed as follows:

$$p\left(A\right) = \begin{cases} b - a, \text{if } a < A < b \\ 0 \text{ else} \end{cases} . \tag{3.13}$$

Considering such prior knowledge allows for compensating noisy observations or incorrect system models. For instance, when x is observed to be greater than or equal to b, since the probability of the true distance A being greater than or equal to b is zero, measures such as ignoring the least squares error in this case can be taken.

Now, let's apply the prior knowledge we have about the parameter A to define the least squares. As A is also a random variable, we can consider not just the likelihood but the joint probability density function (PDF) of $\boldsymbol{x}$ and A, denoted as $p\left(\boldsymbol{x}, A\right)$. Let's denote the optimal value for the distance as $\hat{A}$. The Bayesian mean square error ($Bmse$) is given by the following expression:

$$\begin{aligned} Bmse\left(\hat{A}\right) &= \mathbb{E}\left(A-\hat{A}\right)^2 \\ &= \int\int\left(A-\hat{A}\right)^2 p\left(\boldsymbol{x},A\right) d\boldsymbol{x}dA. \end{aligned} \tag{3.14}$$

Note that this $Bmse$ can be considered as the weighted error summation, different from the traditional Least square error criterion J. Here each error $\left(A-\hat{A}\right)^2$ has its corresponding weight $p\left(\boldsymbol{x},A\right)$ so that the probability of the target error now can be reflected.

Since $p\left(\boldsymbol{x},A\right)=p\left(A|\boldsymbol{x}\right)p\left(\boldsymbol{x}\right)$, we have:

$$Bmse\left(\hat{A}\right) = \int\left(\int\left(A-\hat{A}\right)^2 p\left(A|\boldsymbol{x}\right)dA\right)p\left(\boldsymbol{x}\right)d\boldsymbol{x}. \tag{3.15}$$

Our goal is to have the optimal value $\hat{A}$ from the redefined error criterion $Bmse$. From the axiom of probability $p\left(\boldsymbol{x}\right)\geq 0$, therefore $\hat{A}$ can be obtained by minimizing the integral of A. Taking the derivation we have:

$$\begin{aligned} \frac{\partial}{\partial A}\int\left(A-\hat{A}\right)^2 p\left(A|\boldsymbol{x}\right)dA &= \int\frac{\partial}{\partial A}\left(A-\hat{A}\right)^2 p\left(A|\boldsymbol{x}\right)dA \\ &= \int -2\left(A-\hat{A}\right)p\left(A|\boldsymbol{x}\right)dA \\ &= -2\int Ap\left(A|\boldsymbol{x}\right)dA + 2\hat{A}\underbrace{\int p\left(A|\boldsymbol{x}\right)dA}_{=1}. \end{aligned} \tag{3.16}$$

Set to zero we finally have:

$$\begin{aligned} \hat{A} &= \int Ap\left(A|\boldsymbol{x}\right)dA \\ &= \mathbb{E}\left(A|\boldsymbol{x}\right). \end{aligned} \tag{3.17}$$

In the end, the mean becomes the optimal solution for Bmse. That is, for arbitrary parameter A, the minimum mean square error estimator considering the joint PDF is given by the mean of $p\left(A|\boldsymbol{x}\right)$, denoted as $\mu_{A|\boldsymbol{x}}$. Then what is $p\left(A|\boldsymbol{x}\right)$? How do we find $p\left(A|\boldsymbol{x}\right)$?

Bayes' Rule

$p\left(A|\boldsymbol{x}\right)$ in Eqn. (3.17) is the posterior, the probability of A given $\boldsymbol{x}$, which can be represented using the *likelihood* and the *prior* by the following Bayes' rule:

$$
\begin{aligned}
\underbrace{p(A|\boldsymbol{x})}_{posterior} &= \frac{p(\boldsymbol{x}|A)\,p(A)}{p(\boldsymbol{x})} && \text{(Bayes' Rule)} \\
&= \frac{p(\boldsymbol{x}|A)\,p(A)}{\int p(\boldsymbol{x},A)\,dA} && \text{(marginalization)} \\
&= \frac{p(\boldsymbol{x}|A)\,p(A)}{\int p(\boldsymbol{x}|A)\,p(A)\,dA} && \text{(Def. of Conditional Dist.)} \\
&\propto \underbrace{p(\boldsymbol{x}|A)}_{likelihood}\,\underbrace{p(A)}_{prior} && (p(\boldsymbol{x}) \text{ is } \mathit{Constant}). \qquad (3.18)
\end{aligned}
$$

Note that $\boldsymbol{x}$ are measurements. In other words, $\boldsymbol{x}$ denote the observations of the true value A, which is the parameter of our probability. At this point of time, the observations $\boldsymbol{x}$ are the determined variables observed in the past, even if those variables are sampled from the probability distribution. Therefore $\boldsymbol{x}$ are the known variables thus we say $p(\boldsymbol{x})$ in Eqn. (3.18) becomes constant. In Bayesian estimation, $\boldsymbol{x}$ are also called as *evidence*, since the observations $\boldsymbol{x}$ play a key role of estimating the desired parameter A.

In general, directly obtaining, assuming or formulating the posterior is challenging, compared to the cases of the likelihood (measurement model), or the prior (range of the parameter). Similarly, considering $p(\boldsymbol{x})$ is more challenging, because this distribution is the $generative model$. Therefore, in order to computing the posterior, simply from the likelihood and the prior, we will utilize Eqn. (3.18) in the rest of the chapters in this book.

Consequently, finding the optimal solution of Bayesian mean square error is equal to the solution of MAP, and the solution is determined as the mean of the posterior distribution. In summary, for Bayesian estimation we follow:

- Find the likelihood, mostly representing the observation model
- Assume the reasonable prior
- Compute the posterior from the likelihood and the prior by Bayes' rule
- Use the mean of the posterior as the optimal solution of the Bayesian estimation