Basic statistics – and what you need to know about it

Ola Bengtsson

School of Information Science, Computer and Electrical Engineering, Halmstad University, P O Box 823, Halmstad, Sweden

1 Introduction to statistical representation of a parameter

In the world of today we usually say that we know different things by different amount of certainty, e.g. we can say that a specific kind of battery will last for 5hours ±1hour or that a specific distance (which we measure by a measuring device) is 110cm ±0.5cm, i.e. we do not exactly know the lasting time of the battery or the length of the measured distance. What we do know is that the true lasting time of the battery is somewhere in between 4hours and 6hours (could be any time in between) and that the true distance (the one we are trying to measure) is somewhere in between 109.5cm and 110.5cm, i.e. we have put some bounds on the true values. Instead of expressing the errors with such deterministic bounds (±1hour and ±0.5cm) we can say e.g. that in 67% of the cases the lasting time of the battery is 5hours ±0.5hours and that in 95% of the cases the same lasting time is 5hours ±0.99hours and in 100% of the cases the lasting time is 5hours ±∞ hours. The second representation of the battery lasting time implies that the lasting time can in 5% of the cases either be longer than 6hours or shorter than 4hours. This way of representing a value as a combination of an expected value (for the battery lasting time the expected lasting time is 5hours) and some kind of parameter saying something about how much (and how often) the value differs from the expected value is the basis of statistical representation.

To determine the expected lasting time and how much this time varies it is possible to measure the lasting time of a number of batteries. The measured times are shown in Figure 1 and as we expected most of the lasting times is approximate 5hours. The exact values are given in Table 1.

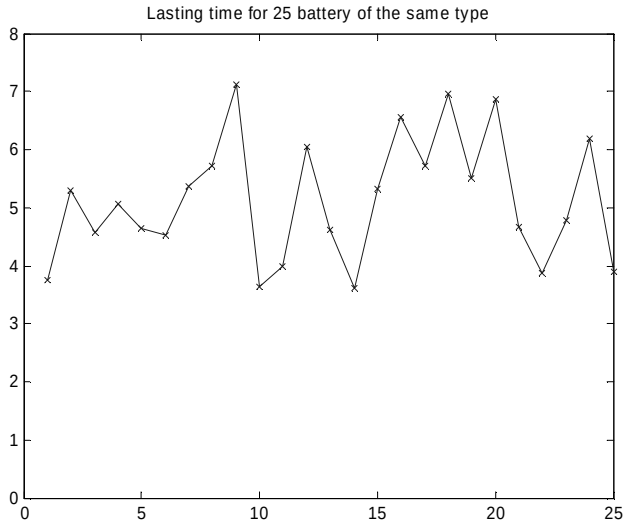

Figure 1: Measured lasting times for 25 batteries of a specific brand.

Table 1: Measured lasting times for 25 batteries of a specific brand [hours].				
3.7656	4.5350	3.9774	6.5532	4.6602
5.2888	5.3710	6.0378	5.7079	3.8602
4.5707	5.7283	4.6102	6.9574	4.7889
5.0558	7.1122	3.6187	5.5045	6.1902
4.6321	3.6427	5.3155	6.8645	3.8838

2 Stochastic variable and its probability distribution

With a stochastic variable (s.v.) we mean a variable that can have any value within some specific bound, e.g. if X is a s.v. and represents the measured distance explained in Section 1 it can have any value in between 109.5cm and 110.5cm. The s.v. follows some kind of probability distributions, $p_X(X = x)$, which tells us the probability of the s.v. to have a specific value. (It exists a lot of different probability distributions, and to learn more about them and their statistical properties I strongly recommend you to study some course in basic statistics). In mobile robotics, though, one distribution in particular is used more frequently than all the others together, and that is the Normal (also called Gaussian) distribution. Another often used distribution is the Uniform distribution, which is a distribution that gives the same probability to all values within specific bounds, i.e. if a s.v., X, is uniformly distributed with an expected value of 5 ±2 it is equally likely that X = 4 as X = 5 or X = 7 etc. Anyone that has ever used the rand(…) function in e.g. ANSI-C has actually used an uniformly distributed stochastic variable.

2.1 Expected value and variance

The expected value of the s.v. is the weighted mean of all possible values for the s.v., i.e. in the case where s.v. is continously the expected value becomes as in Equation 2.1a. The case where the s.v. is instead discrete is shown in Equation 2.1b.

$$E[X] = \int_{-\infty}^{\infty} x.f_X(x)dx \tag{2.1a}$$

$$E[X] = \sum_{k=-\infty}^{\infty} k.p_X(k) \tag{2.1b}$$

The variance, i.e. how much X varies relative the expected value, of the same s.v. is shown in Equation 2.2a and Equation 2.2b for continously respectice discrete s.v. X. The variance of a s.v. X is commonly referred to as σ_X^2, with σ_X referred to as the standard deviation of the same variable.

$$V[X] = \sigma_X^2 = \int_{-\infty}^{\infty} (x - E[X])^2.f_X(x)dx \tag{2.2a}$$

$$V[X] = \sigma_X^2 = \sum_{K=-\infty}^{\infty} (k - E[X])^2.p_X(k) \tag{2.2b}$$

These values are absolutely crusial to understand, as the entire statistical framework is build around these values. When we e.g. tries to estimate the position of a mobile robot we usually describe the position as an expected value (which is the

position where we think the robot is located) and a variance that gives the information of how certain the estimated position is. A small variance means that the robot is very sure of its position and a large variance means the robot is uncertain of its position. Actually, if the variance would be infinite that would mean that the robot could be located anywhere in the world.

3 Gaussian (Normal) distribution

The probability function of a gaussian distributed s.v. is given in Equation 3.1, which is – as one would expect – a function of the expected value and the variance, which means that if we know that a s.v. is gaussian distributed then we only need the expected value and the variance to describe the entire probability distribution.

$$p_X(x) = \frac{1}{\sigma_X . \sqrt{2\pi}} e^{-\frac{(x-E[X])^2}{2\sigma_X^2}} \tag{3.1}$$

To say that a s.v., X, is gaussian distributed we prefer to say that X follows the gaussian distribution with the expected value and the standard deviation, which is the square root of the variance. (From now on and further we also call the expected value the mean value of X and denote it as m_X.) This is written as in Equation 3.2 and seen in Figure 2.

$$X \sim N(m_X, \sigma_X) \tag{3.2}$$

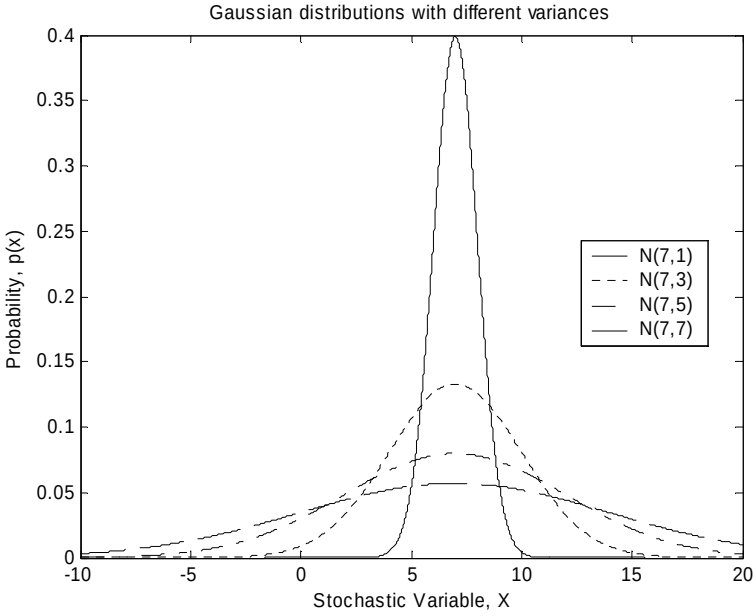

Figure 2: Stochastic variable following a Gaussian distribution.

Question 1: What does it mean if the robot tells you it has moved X1 meters with X1 being a s.v. with mean 7 and variance 1 compared to if the robot should have moved X2 meters with X2 being a s.v. with mean 7 and variance 0.25? If you were to control the robot, which one would you prefer?

Question 2: Does the Gaussian distribution limits the possible values of the s.v., i.e. in Question 1 – is it possible that the robot could actually be several 1000's meters away from the mean value?

Question 3: What does it mean if a specification of a sensor says that the sensor measures a distance in the range 1 – 50 meters and has an error that is normally distributed N(0, σ) with σ = 0.1 meters.

Remark: The standard deviation tells you that ~68.3% of your values will stay within the bound $[m_X - \sigma_X, m_X + \sigma_X]$, ~95% of the values in between $[m_X - 2\sigma_X, m_X + 2\sigma_X]$, ~99% of the values in between $[m_X - 3\sigma_X, m_X + 3\sigma_X]$ etc. (Take a look in a table of the Gaussian distribution.)

3.1 Estimating the distribution parameters of a Gaussian distribution from observations

By making a lot of observations we can estimate the distribution parameters, i.e. the expected (mean) value and the variance. Of course, the first thing to do is to verify that the measurement errors really follow a Gaussian distribution. If notm then we should of course use another distribution. Consider the example with the measured lasting times of the batteries in Section 1. By plotting a histogram of the measurements it is possible to verify whether this process is Gaussian or not. The histogram on the basis of the 25 measurements is shown in Figure 3.

The distribution in Figure 3 does not look even close to the Gaussian distributions shown in Figure 2 and one might (wrongly) assume that the errors are not Gaussian distributed. The problem here is the lack of observations, i.e. 25 observations are not enough, why Figure 4 shows the same distribution but is instead based on 100 observations.

Now the values seem to be Gaussian distributed and we can estimate the distribution parameters, i.e. the expected (mean) value together with the variance. A good (and statistical correct) way of estimating the expected value is to use the aritmetic mean value, i.e. according to Equation 3.3. (The ^ means that this parameter is estimated, i.e. it is not the true value!)

$$\hat{m}_X = \frac{1}{N}\sum_{k=1}^{N} X(k) \tag{3.3}$$

A good estimate of the variance (and also statistically correct – to know why the formula looks exactly as it does you should study a course in basic statistic) is the average of the sum of squared distances between the measurements and the expected value, i.e. according to Equation 3.4 (Remark that the variance should be divided by N-1 and not N as was the case with the expected value.)

$$\hat{\sigma}_X^2 = \frac{1}{N-1}\sum_{k=1}^{N} (X(k) - \hat{m}_X)^2 \tag{3.4}$$

This means that the battery lasting time, X, is Gaussian distributed according to Equation 3.5.

$$X \sim N\left(\hat{m}_X, \hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2}\right) \quad \text{on the basis of 100 observations } \hat{m}_X = 5.1259 \text{ and } \hat{\sigma}_X = 1.0669 \tag{3.5}$$

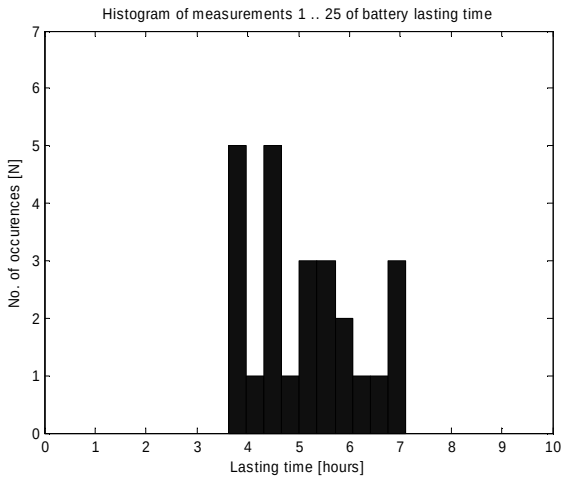

Figure 3: Histogram of battery lasting time (using 25 observations).

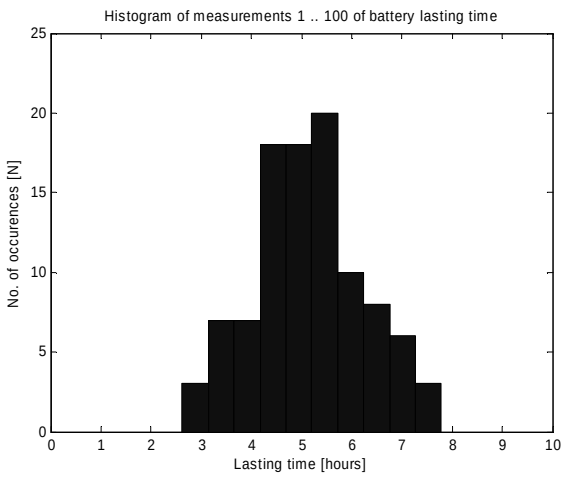

Figure 4: Histogram of battery lasting time based on 100 observations.

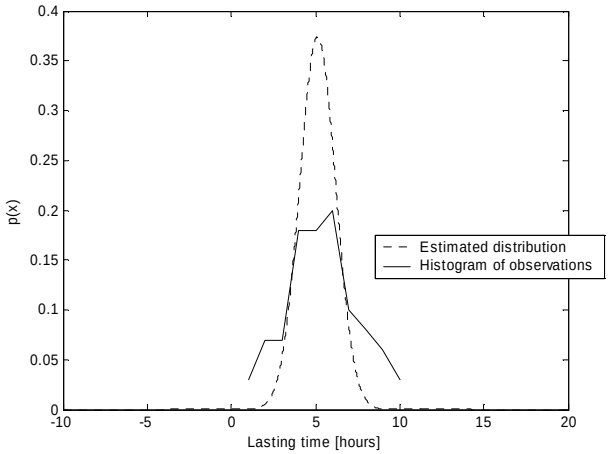

Figure 5: Comparing the observations to the estimated Gaussian distribution.

4 Linear combinations of normally distributed independent s.v.

It is of course possible to combine two (or more) s.v. to form a new, third, s.v., e.g. if we have the two s.v. $X_1 \sim N(m_1, \sigma_1)$ and $X_2 \sim N(m_2, \sigma_2)$ then we can create a new s.v. as Y = X_1 + X_2, i.e. Y is a linear combination of X_1 and X_2. One of the really nice properties of the Gaussian distribution is that if we linearily combine several s.v. then also the result will follow a Gaussian distribution. Actually Y will follow a Gaussian distribution with parameters according to Equation 4.1.

$$Y \sim N\left(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right) \tag{4.1}$$

The parameters are easily calculated using the calculation rules for s.v. Some very important rules are given in Equation 4.2 – 4.5.

$$E[aX + b] = aE[X] + b \tag{4.2}$$

$$V[aX + b] = a^2 V[X] \tag{4.3}$$

$$E[X_1 + X_2] = E[X_1] + E[X_2] \tag{4.4}$$

$$V[X_1 + X_2] = V[X_1] + V[X_2] \tag{4.5}$$

Question: Assume we measure the distance of something and we know that our measurements, $\hat{D} \sim N(D, \sigma_D)$, are Normally distributed with some known variance. If we make 5 measurements on the same object and use the average value as our estimate of D, i.e. we want to have the s.v. Y according to Equation 4.6.

$$Y = \tfrac{1}{5}D + \tfrac{1}{5}D + \tfrac{1}{5}D + \tfrac{1}{5}D + \tfrac{1}{5}D = \tfrac{1}{N}\sum_{i=1}^{N=5} D_i \qquad (4.6)$$

What is the expected value of Y? What will the variance of Y be? What will the standard deviation of Y be? How many measurements would you need to have to use to get a standard deviation of $0.1\sigma_D$? The fact that the variance decreases by taking the average of several measurements is a very good property – if you have a sensor with a large variance how would you use the property of the variance in a good way?

Question 5: If we measure some distance, D, by using two different methods ($\hat{D}_1$ and $\hat{D}_2$) which give both different expectation values and different variances. Assume that $\hat{D}_1 \sim N(6.5,1)$ and $\hat{D}_2 \sim N(5.78,10)$ – how would you use the measurements? Assume we want the combination that gives the smallest variance possible, does such an optimal combination of the measurements exist? If yes – how big is the variance? Is the variance bigger, the same or smaller than the individual variances?

5 Non-linear combinations of normally distributed independent s.v.

Assume we measure the relative movement of an airplane (see Figure 6) by a distance, $\hat{d}_i = d_i + \varepsilon_d$, which is the true distance plus some noise, which is normally distributed with zero mean and known variance σ_d^2. We also measures (by a gyro or something similar) the change in heading angle of the airplane, $\hat{\alpha}_i = \alpha_i + \varepsilon_\alpha$, i.e. the change in the airplane's heading relative ground. Also the error in the angle measurement is normally distributed with zero mean and known variance σ_α^2. As the two parameters are given by different measuring systems we can further assume that the errors are uncorrelated, i.e. the co-variance between ε_d and ε_α is zero. (The co-variance is defined in Section 6.)

From Section 4 we know that we easily can estimate the, by the airplane, total distance travelled at time step N and the variance at time step N by simply making a new s.v. according to Equation 5.1, which will have the expected value according to Equation 5.2. and variance according to Equation 5.3. This is straight forward as we know that if several Gaussian distributed s.v. are linearly combined then also the combined s.v. will be Gaussian distributed.

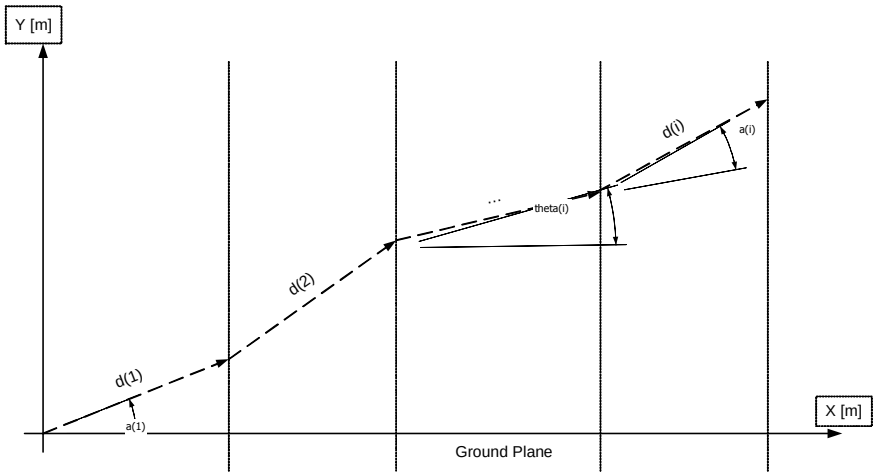

Figure 6: Airplane flying according to some specific route.

$$D = d_1 + d_2 + ... + d_N = \sum_{k=1}^{N} d(k) \tag{5.1}$$

$$E\left[\hat{D}\right] = E\left[\sum_{k=1}^{N}\{d(k) + \varepsilon_d\}\right] = \sum_{k=1}^{N} E\left[d(k)\right] + \sum_{k=1}^{N} E\left[\varepsilon_d\right] = \sum_{k=1}^{N} E\left[d(k)\right] + \sum_{i=1}^{N} 0 = \sum_{i=1}^{N} d(k) \tag{5.2}$$

$$V\left[\hat{D}\right] = V\left[\sum_{k=1}^{N}\{d(k) + \varepsilon_d\}\right] = \sum_{k=1}^{N} V\left[d(k)\right] + \sum_{k=1}^{N} V\left[\varepsilon_d\right] = \sum_{k=1}^{N} 0 + \sum_{k=1}^{N} V\left[\varepsilon_d\right] = N\sigma_d^2 \tag{5.3}$$

Question 6: Assume the plane flies for a really long time, i.e. $N \to \infty$, what would this mean to $\hat{D}$, i.e. what happens to the expected value and what happens to the variance? Do we know how far the plane has gone?

Also the heading angle θ of the airplane is easily estimated as this parameter is the sum of the relative changes in rotation, i.e. according to Equation 5.4, which has an expected value according to Equation 5.5 and variance according to Equation 5.6.

$$\theta = \theta_0 + \alpha(0) + \alpha(1) + ... + \alpha(N-1) = \theta_0 + \sum_{k=0}^{N-1} \alpha(k) \tag{5.4}$$

$$E\left[\hat{\theta}\right] = E\left[(\theta(0) + \varepsilon_\theta(0)) + \sum_{k=0}^{N-1}\{\alpha(k) + \varepsilon_\alpha(k)\}\right] = \theta(0) + \sum_{k=0}^{N-1} \alpha(k) \tag{5.5}$$

$$V\left[\hat{\theta}\right] = V\left[(\theta(0) + \varepsilon_\theta(0)) + \sum_{k=0}^{N-1}\{\alpha(k) + \varepsilon_\alpha(k)\}\right] = V\left[\varepsilon_\theta(0)\right] + \sum_{k=0}^{N-1} V\left[\varepsilon_\alpha(k)\right] \tag{5.6}$$

If we instead want to know how far relative the starting position (parallel to the ground) the airplane has moved, i.e. if we instead want to estimate X from our measurements of d and α, then we can write X at time N as the previous value of X plus the latest contribution of d and α, i.e. according to Equation 5.7.

$$X_N = X_{N-1} + d_{N-1}\cos(\theta_{N-1} + \alpha_{N-1}) \tag{5.7}$$

The estimate $\hat{X}_N$ of X_N is given in Equation 5.8.

$$\hat{X}_N = \hat{X}_{N-1} + \hat{d}_{N-1}\cos(\hat{\theta}_{N-1} + \hat{\alpha}_{N-1}) = (X_{N-1} + \varepsilon_{X_{N-1}}) + (d_{N-1} + \varepsilon_d)\cos(\theta_{N-1} + \varepsilon_{\theta_{N-1}} + \alpha_{N-1} + \varepsilon_\alpha) \tag{5.8}$$

It is important to notice that Equation 5.5 is non-linear because it contains the term $\cos(\theta_{N-1} + \varepsilon_{\theta_{N-1}} + \alpha_i + \varepsilon_\alpha)$, which means for X_N to be kept Gaussian we need to linearize Equation 5.8 around $\theta_{N-1} + \alpha_{N-1}$. (It is also important to notice that Equation 5.8 would have been linear if we had known $\theta_{N-1} + \alpha_{N-1}$ perfectly, i.e. if we would have been able to measure it without making any error, which is, of course, not possible because all measurement systems make some kind of errors.) In Equation 5.9 we use a first order Taylor Expansion to linearize Equation 5.8 around $\theta_{N-1} + \alpha_{N-1}$.

$$\hat{X}_N = (X_{N-1} + \varepsilon_{X_{N-1}}) + (d_{N-1} + \varepsilon_d)\cos(\theta_{N-1} + \varepsilon_{\theta_{N-1}} + \alpha_{N-1} + \varepsilon_\alpha) \approx$$

$$\approx (X_{N-1} + \varepsilon_{X_{N-1}}) + (d_{N-1} + \varepsilon_d)\left(\cos(\theta_{N-1} + \alpha_{N-1}) - (\varepsilon_{\theta_{N-1}} + \varepsilon_\alpha)\sin(\theta_{N-1} + \alpha_{N-1})\right) \quad (5.8)$$

We now have $\hat{X}_N$ as a linear function of d and α and we can calculate the expected value and the variance in the same way as before, i.e. the expected value becomes as shown in Equation 5.9.

$$E\left[\hat{X}_N\right] \approx E\left[(X_{N-1} + \varepsilon_{X_{N-1}}) + (d_{N-1} + \varepsilon_d)\left(\cos(\theta_{N-1} + \alpha_{N-1}) - (\varepsilon_{\theta_{N-1}} + \varepsilon_\alpha)\sin(\theta_{N-1} + \alpha_{N-1})\right)\right] =$$

$$= E\left[X_{N-1}\right] + E\left[d_{N-1}\cos(\theta_{N-1} + \alpha_{N-1})\right] = X_{N-1} + d_{N-1}\cos(\theta_{N-1} + \alpha_{N-1}) \quad (5.9)$$

Only a few parts of Equation 5.9 contribute to the expected value because the expected value of ε_d, ε_α and $\varepsilon_{X_{N-1}}$ are all zero. The variance becomes as shown in Equation 5.10 (for simplicity we set $\phi = \theta_{N-1} + \alpha_{N-1}$).

$$V\left[\hat{X}_N\right] \approx V\left[(X_{N-1} + \varepsilon_{X_{N-1}}) + (d_{N-1} + \varepsilon_d)\left(\cos(\phi) - (\varepsilon_{\theta_{N-1}} + \varepsilon_\alpha)\sin(\phi)\right)\right] =$$

$$= V\left[X_{N-1}\right] + V\left[\varepsilon_{X_{N-1}}\right] + V\left[(d_{N-1} + \varepsilon_d)\left(\cos(\phi) - (\varepsilon_{\theta_{N-1}} + \varepsilon_\alpha)\sin(\phi)\right)\right] = \quad (5.10)$$

$$= \sigma^2_{X_{N-1}} + \left(-d_{N-1}\sin(\phi)\right)^2\sigma^2_\alpha + \left(\cos(\phi)\right)^2\sigma^2_d + \left(-d_{N-1}\sin(\phi)\right)^2\sigma^2_{\theta_{N-1}}$$

Two things are very important to notice, first that the linearization only affects the prediction of the variance and second that Equation 5.8 ends up as the partial derivatives of $\hat{X}_N = \hat{X}_{N-1} + \hat{d}_{N-1}\cos(\hat{\theta}_{N-1} + \hat{\alpha}_{N-1})$ with respect to our uncertain parameters raised to the power of two multiplied with the variance of respective parameter, which is very nice as this can be written as the law of error propagation (seen in Equation 5.11). The partial derivatives are given in Equation 5.12. (In many papers it says that the update equations of the robots position are linearized around its previous state and the linearization done in Equation 5.8 is what they mean by it.) If we take a really close look at Equation 5.10 we also notice that the variance of $\hat{X}_N$ is the variance $\hat{X}_{N-1}$ plus some errors introduced by the latest movement, i.e. errors introduced from d_{N-1} and α_{N-1}, i.e. the variance is our old variance and some contributions caused by our latest movement.

$$V\left[\hat{X}_N\right] = \left(\frac{\partial \hat{X}_N}{\partial X_{N-1}}\right)^2 \sigma^2_{X_{N-1}} + \left(\frac{\partial \hat{X}_N}{\partial d_{N-1}}\right)^2 \sigma^2_d + \left(\frac{\partial \hat{X}_N}{\partial \alpha_{N-1}}\right)^2 \sigma^2_\alpha + \left(\frac{\partial \hat{X}_N}{\partial \theta_{N-1}}\right)^2 \sigma^2_{\theta_{N-1}} \quad (5.11)$$

$$\frac{\partial \hat{X}_N}{\partial X_{N-1}} = 1 \qquad \frac{\partial \hat{X}_N}{\partial d_{N-1}} = \cos(\phi) \qquad \frac{\partial \hat{X}_N}{\partial \alpha_{N-1}} = -\hat{d}_{N-1}\sin(\phi) \qquad \frac{\partial \hat{X}_N}{\partial \theta_{N-1}} = -\hat{d}_{N-1}\sin(\phi) \quad (5.12)$$

The estimate of X_N and the variance are shown, for N = 1, …, 20, in Figure 7.

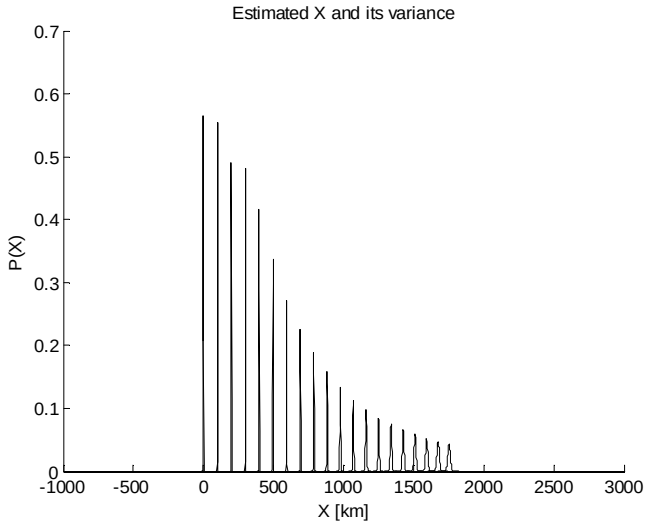

Figure 7: Plot of estimated X and the corresponding variance.

Question 7: Derive the expected value and the variance for $\hat{Y}_N$, i.e. how far up in the sky the airplane is at time step N and then use the law of error propagation – do you get the same result? If the airplane starts in origin and flies with a constant change in the heading angle of +2deg and with a constant speed, i.e. all d(k) are the same, where will the airplane be (in X and Y) at time step N = 20? Which of the parameters ($\hat{X}_N$ or $\hat{Y}_N$) is estimated the best, i.e. which of the parameters has the smallest variance? If they are not the same, why are they not the same?

6 Multidimensional Gaussian (Normal) distribution

When we talk about robot systems, and positioning of such systems, we often need to estimate parameters in several dimensions, e.g. a robot's position given as x, y and z co-ordinates or as x, y and θ where x and y represent the position on the ground, z the altitude and θ the heading. All these parameters have different uncertainty (different large variances) and they can be either uncorrelated or correlated. When it comes to estimate the position of a robot the parameters are most often correlated. For this reason it is crucial to also know how to deal with the Gaussian distribution in several dimensions, which is not more difficult than dealing with Gaussian distributions in one dimension. The difference is that instead of using scalars we have to use matrices. Equation 6.1 gives the Gaussian distribution for a multidimensional distribution.

$$p_X(x) = \left(\frac{1}{(2\pi)^N |\sum_X|} \right)^{\frac{1}{2}} e^{-\frac{1}{2}(x-m_X)^T \sum_X^{-1} (x-m_X)} \tag{6.1}$$

In Equation 6.1 $\sum_X$ is the covariance matrix (symmetric and of size NxN), X is a column vector (of size Nx1) and m_X is the mean vector, or the vector of expected values, (also of the size Nx1). The co-variance matrix contains the variances of each dimension (which are the diagonal values) and the co-variances in between all combinations of the dimensions. If we assume that X has three dimensions then the co-variance matrix will look like in Equation 6.2.

$$\Sigma_X = \begin{pmatrix} \sigma_{x1}^2 & C(x1,x2) & C(x1,x3) \\ C(x2,x1) & \sigma_{x2}^2 & C(x2,x3) \\ C(x3,x1) & C(x3,x2) & \sigma_{x3}^2 \end{pmatrix} \tag{6.2}$$

Also known is that $C(x_i,x_j) = C(x_j,x_i)$ and therefore the co-variance matrix is always symmetric. The co-variance matrix is also always positive definite, i.e. always produces positive eigenvalues. The co-variance between two s.v. is defined according to Equation 6.3. For a continous s.v. the co-variance is calculated as in Equation 6.4a and for discrete s.v. as in Equation 6.4b.

$$C(X,Y) = E[(X - m_X)(Y - m_Y)] \tag{6.3}$$

$$C(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - m_X)(y - m_Y) f_{X,Y}(x,y)dxdy \tag{6.4a}$$

$$C(X,Y) = \sum_j \sum_k (j - m_X)(k - m_Y) p_{X,Y}(j,k) \tag{6.4b}$$

It is important to say that if C(X,Y) = 0, then X and Y are uncorrelated, i.e. they have no influence on each other. To see what happens we will, once more, do the example with the flying airplane (as we did in Section 5). We can write the position of the airplane at time k as X(k), which is a vector consisting of an x, y and θ component. The predicted position at time k+1, i.e. X(k+1), then becomes a function of the position at time k, the relative change in movement d(k) and the relative change in heading α(k), i.e. X(k+1) can be written as in Equation 6.5. (We gather all the input parameters, i.e. d and α, in the input vector U.)

$$X(k+1) = f(X(k),U(k)) = \begin{pmatrix} x(k) + d(k)\cos(\theta(k) + \alpha(k)) \\ y(k) + d(k)\sin(\theta(k) + \alpha(k)) \\ \theta(k) + \alpha(k) \end{pmatrix} \tag{6.5}$$

The co-variance matrix at time step k+1 is given by the law of error propagation (applied on matrices) and shown in Equation 6.6.

$$P(k+1|k) = \nabla f_X P(k|k) \nabla f_X^T + \nabla f_U U(k+1) \nabla f_U^T \tag{6.6}$$

In Equation 6.6 ∇f_X and ∇f_X are the Jacobian matrices with respect to the state variables, X, and the input variables, U, respectively, i.e. become as in Equations 6.7 and 6.8.

$$\nabla f_X = \begin{pmatrix} 1 & 0 & -d(k)\sin(\theta(k) + \alpha(k)) \\ 0 & 1 & d(k)\cos(\theta(k) + \alpha(k)) \\ 0 & 0 & 1 \end{pmatrix} \tag{6.7}$$

$$\nabla f_U = \begin{pmatrix} \cos(\theta(k)+\alpha(k)) & -d(k)\sin(\theta(k)+\alpha(k)) \\ \sin(\theta(k)+\alpha(k)) & d(k)\cos(\theta(k)+\alpha(k)) \\ 0 & 1 \end{pmatrix} \tag{6.8}$$

The predicted co-variance matrices for time step 1, …, 20 are shown in Figure 8. (Only the variances (or actually the standard deviations) for X and Y are shown as it is rather difficult to show more than two dimensions in the same plot.)

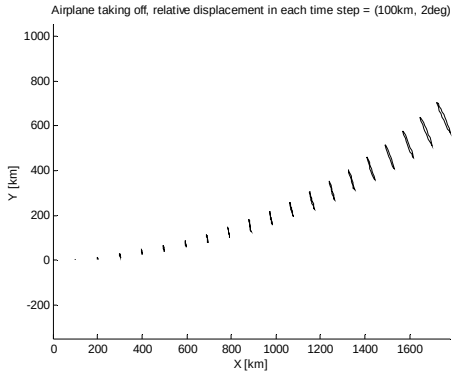

Figure 8: Airplane taking off.

As can be seen in Figure 8, the variance in y changes more drastically than the variance in x, which actually means that from our measurements we can estimate the position in x with better certainty than in y. This is correct because the error in α(k) affects the error in y much more than the error in x. Another thing that should be noticed is that the errors in x and y are getting more correlated with time, which depends on the airplane's heading angle.