Statistical Orbit Determination: Appendix A

Objective: Mastering Statistical Orbit Determination

TO:		**Date:**	16 Apr 2018
CC:		**Memo:**	SOD: Notes-Appendix A
		From:	Christopher R. Simpson
REF:		**Email:**	simpsonchristo@gmail.com

SUMMARY:

OVERVIEW:

A random variable X is a real-valued function associated with the chance outcome (observation of result) of an experiment. Each possible observation is (in our notation) denoted x. All possible values x_i are the domain of X.

What types of random variables are there?

There are two types, discrete and continuous. Think of discrete like a list of numbers or the order runners finished in a race. The possible outcomes of X are a set of finite real numbers. Continuous is an interval on a real line like the length of a student's desk in inches. The probability at a point is 0 (because the area under the curve is 0 at a point).

What are the rules governing probability?

To answer this we turn to the axioms of probability. If S is the sample space (set of all possible outcomes) then let A be the subset of points of the set S (i.e. a collection of one or more possible trials, like a sample). Consider a die; $s = 1, 2, 3, 4, 5, 6$ and a subset may be $a = 3$. The numerical probability of event A occurring is denoted $p(A)$. Then we can consider these three axioms which will govern our further attempts at illuminating the use of probability in determining orbits.

1. $p(A) \geq 0$
2. $p(S) = 1$
3. $p(A + B) = p(A) + p(B)$; provided A and B are mutually exclusive, $P(AB) = 0$.

What if A and B are not mutually exclusive?

$$p(A + B) = p(A) + p(B) - p(AB)$$

What is the complementary rule?

The complementary rule compliments $p(A)$. The chance of $p(A)$ not occurring is given by $p(\bar{A}) = 1 - p(A)$.

And? Or? What do they mean?

Consider a Venn diagram. Where circle A and circle B intersect is the probability that A and B occur, $p(AB)$. This is denoted as $A \cap B = AB$. The possibility of A or B occurring is the union of the two Venn diagrams for the whole sample space, $p(A + B)$. This is denoted as $A \cup B = A + B$.

Okay, I get how to get if A or B occurs but what about both occurring?

In order to get to $p(AB)$ from $p(A)$ and $p(B)$ we need to know if $p(A)$ and $p(B)$ are independent. Two events are independent if the probability that an event occurs is the same regardless if the first event occurred. $p\left({}^{A}/_{B}\right) = p(A)$ is the required condition for independence $p(AB) = p(A)p(B)$. If the events are not independent then; $p(A/B) = p(AB)/p(B)$.

You said the probability at a point is 0 for a continuous random variable. How do I find the probability for a continuous random variable, like say the measured length of my desk?

For a continuous random variable we assume that all events of practical interest will be represented by intervals on the real line. A probability density function, $f(x)$, which represents the probability of X assuming a value on the interval $(x, x + dx)$.

$$p(x \leq X \leq x + dx) = f(x)dx$$

The rules we mentioned earlier, that the probability of a subset is equal to or greater than 0, the probability of the entire sample space is 1, and the probability of event A or B is equal to their summed individual probabilities if they're mutually exclusive: How do they matter here?

Well clearly we need to update our definitions to include the idea of the probability density function $f(x)$. These axioms haven't changed, just their representation:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $p(a \leq X \leq c) = \int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$

For 3 just think of it as increasing the interval.

Great, the probability density function gives us the probability based on the clumping of the data. Well does it describe the distribution of the continuous random variable X?

No! The probability density function just measures the "density," or clumping of the data to give us a probability estimate. The distribution function, $F(x)$ describes the actual distribution. $\frac{dF(x)}{dx} = f(x)$ so that if we are interested in the event $X \leq x$ then $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$. It follows that $F(-\infty) = 0$ and $F(\infty) = 1$. We can confirm this using axioms 1 and 2; $0 \leq F(x) \leq 1$. $F(x)$ is also monotonically increasing.

How is that useful?

Well the distribution function is the actual result returned when we evaluate the probability density function like in the third axiom. For example, $p(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$.

Okay. What are some useful ways to talk about the probability distributions?

Let's start with an easy one. The <u>expected value</u> or the <u>mean</u> of X is $E(X)$. We find $E(X) = \int_{-\infty}^{\infty} xf(x)dx$.

A more general way to think about this is if we consider a second random variable Y that is a function of X so $Y = g(X)$. Then $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$.

So the mean is just the expected value of the continuous random variable. What about variance or standard deviation?

Going back to our explanation of the expected value: $E(X)$ is sometimes written as λ_1 or λ and is called the arithmetic mean of X. Geometrically, λ_1 is one of a number of possible devices for locating the centroid (center) of the probability distribution with respect to the origin.

The kth moment of X about the origin is $E[X^k] = \lambda_k = \int_{-\infty}^{\infty} x^k f(x)dx$. The kth moment of X about the mean λ_1 is $\mu_k = E(X - \lambda_1)^k = \int_{-\infty}^{\infty} (x - \lambda_1)^k f(x)dx$. For $k = 1$, $\mu_1 = 0$. The variance is difference from mean, right? So for $k = 2$, $\mu_2 = \sigma^2$, or the variance of X or the second moment of X about the mean. It is one measure of the dispersion of the distribution about its mean value.

Why is this important?

In guidance and estimation applications we are most concerned with the first two moments, namely the mean and variance. Future values for the state of the dynamic system can be obtained by propagating the joint density function $f(u, v)$ forward in time and using it to calculate mean and variance. The equations for propagating the mean and variance of a random vector X are discussed in Section 4.8 of the text.

What is a Moment Generating Function?

A moment generating function is an expected value function that generates the moments $\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_n$ of the random variable X. In other words, $g(X) = e^{\theta x}, E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$. Since $e^{\theta x} = 1 + \theta x + \frac{(\theta x)^2}{2!} + \cdots + \frac{(\theta x)^n}{n!}$ we see that $E(e^{\theta x}) = \lambda_0 + \theta\lambda_1 + \frac{\theta^2\lambda_2}{2!} + \cdots \frac{\theta^n\lambda_n}{n!}$. The moment generating function $E(e^{\theta x}) = M_X(\theta)$ we can use to find moments about the origin by $\frac{\partial^k M_X(\theta)}{\partial\theta^k}|_{\theta=0} = \lambda_k$. We can replace x with any function $h(X)$to find the moment generating function for that function. So if we wanted to find the moments about the mean, $h(X) = X - \lambda_1$ then $M_{X-\lambda_1}(\theta) = e^{-\theta\lambda_1}M_X(\theta)$.

What shape does the probability density function take?

The easy answer is whatever the slope of the distribution function $F(X)$ takes. Really, there are two important ones for this book: The uniform (rectangular) and the normal (Gaussian) distribution.

The uniform distribution occurs when X has equal probability over the range $a \leq X \leq b$. We can write the density function as $f(x) = \frac{1}{b-a}, a \leq x \leq b$ and 0 elsewhere. The first two moments: $E(X) = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a}dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{a+b}{2}$ and $\sigma^2 = \frac{(b-a)^2}{12} = \int_a^b [x - E(X)]^2 f(x)dx$. The reason we care about the uniform distribution is its utility. We can elicit properties of a continuous probability distribution by converting first to a uniform distribution and then into a given continuous distribution.

The Gaussian distribution is the most commonly found distribution found as $f(x) = \frac{1}{\sqrt{2\pi}b}\exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right]$ over $-\infty \leq x \leq \infty$ and $b > 0$. The moment generating function $M_X(\theta) = \exp\left[\frac{\theta^2 b^2}{2} + a\theta\right]$. $\lambda_1 = a$, $\sigma^2 = b^2$. Confidence intervals include for $1\sigma, 2\sigma$, and 3σ, $p = 0.68268, 0.95449, 0.99730$, respectively.

Okay, but I want to estimate position and velocity. How do I get there?

We're getting closer. First we need to cover two random variables. We've covered the necessary background for one, now let's see how two interact.

$$F(x_o, y_o) = p\{X \leq x_o, Y \leq y_o\}$$

Returning to our distribution function properties:

For all x, y the distribution function $0 \leq F(x,y) \leq 1$. Like before, $F(-\infty, y) = F(x, -\infty) = 0$ and $F(\infty,\infty) = 1$. But when $F(\infty, y) = p(Y \leq y)$ and $F(x, \infty) = p(X \leq x)$.

The distribution function is the integral of the area element and the density function. $f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$ or $p(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)dx\, dy$.

What if I want to determine the probability of one variable given the joint probability of the two?

$p(X \leq x$,no condition on $Y) = F(x, \infty)$ and for the continuous case this translates to $g(x) = \int_{-\infty}^{\infty} f(x,y)dy$ or $h(y) = \int_{-\infty}^{\infty} f(x,y)dx$. Integrate over the unwanted variable to find the marginal density function of a random variable.

How do I know if they're independent?

$$f(x,y) = g(x)h(y)$$

Let's get back to those expected value functions. What's the mean or variance of two continuous random variables?

$E[\phi(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(x,y)f(x,y)\, dx\, dy$ gives the expected value of an arbitrary function of the two continuous random variables. If we set $\phi(X,Y) = X^l Y^m$ the expected value is updated to $E[X^l Y^m] =$

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^l y^m f(x,y)dxdy$. This means (referring to the expected values for one variable) $E[X^l Y^m] = \lambda_{lm}$ or the lmth moment of X, Y about the origin. To find the lmth moment about the mean, $\phi(X,Y) = [X - \lambda_{10}]^l [Y - \lambda_{01}]^m$. This returns $\mu_{lm} = E\{[X - \lambda_{10}]^l [Y - \lambda_{01}]^m\}$.

Particular cases of μ_{lm} and λ_{lm} give:

l	m		
0	0	$\lambda_{00} = 1$	
1	0	$\lambda_{10} = E(X)$	Mean of X
0	1	$\lambda_{01} = E(Y)$	Mean of Y
0	0	$\mu_{00} = 1$	
1	1	$\mu_{11} = E\{[X - E(X)][Y - E(X)]\}$	Covariance of X and Y
2	0	$\mu_{20} = \sigma^2(X)$	Variance of X
0	2	$\mu_{02} = \sigma^2(Y)$	Variance of Y

How would I calculate the covariance?

An example, similar to computing the variance for 1 variable:

$$\mu_{11} = E\{[X - E(X)][Y - E(X)]\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \lambda_{10})^1 (y - \lambda_{01})^1 f(x,y)dxdy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (xy - \lambda_{10})(y - \lambda_{01}) f(x,y)dxdy$$

Great! Now let's turn that into something I can use for state matrices like a matrix.

Look at this handy-dandy variance-covariance matrix. The symmetric matrix:

$$P = E\left[\begin{bmatrix} X - E(X) \\ Y - E(Y) \end{bmatrix} \begin{bmatrix} X - E(X) & Y - E(Y) \end{bmatrix}\right]$$
$$P = E\begin{bmatrix} \big(X - E(X)\big)^2 & \big(X - E(X)\big)\big(Y - E(Y)\big) \\ \big(X - E(X)\big)\big(Y - E(Y)\big) & \big(Y - E(Y)\big)^2 \end{bmatrix}$$
$$P = \begin{bmatrix} \sigma^2(X) & \mu_{11} \\ \mu_{11} & \sigma^2(Y) \end{bmatrix}$$

The covariance of two random variables is often written in terms of a correlation coefficient, ρ_{XY}. This coefficient helps us determine the correlation between X and Y. -1 or 1, negative or positive correlation, respectively. Near 0 and there is no correlation.

$$\rho_{XY} = \frac{\mu_{11}}{\sigma(X)\sigma(Y)}$$

The correlation coefficient is also the covariance between standardized random variables.

What does a posteriori and a priori mean?

A posteriori means given the observation vector we estimate the state vector. *A priori* we estimate the state vector prior to the observation vector.