



**UNIVERSIDAD MICHOACANA DE
SAN NICOLÁS DE HIDALGO**



FACULTAD DE INGENIERÍA ELÉCTRICA

**EL FILTRO DE KALMAN Y SU APLICACIÓN A LA MEDICIÓN DE
INCLINACIÓN MEDIANTE UN ACELERÓMETRO Y UN
GIROSCOPIO ELECTRÓNICOS**

TESIS

**QUE PARA OBTENER EL TÍTULO DE:
INGENIERO EN ELECTRÓNICA**

**PRESENTA:
GERARDO RAMÍREZ ARREDONDO**

**ASESOR:
Dr. JOSÉ JUAN RINCÓN PASAYE**

MORELIA MICHOACÁN, JULIO DEL 2013

Agradecimientos

A mi hermano, que donde quiera que esté, sé que siempre me cuida y me levanta cuando no tengo fuerza para hacerlo por mi mismo.

A mis padres, por haberme apoyado en los buenos tiempos y en los no tan buenos, por siempre creer en mí, aún cuando yo no lo hacía.

A mi familia, que nunca pierde su confianza en mí.

A mis amigos que son la familia que yo elegí.

A todos esos seres queridos, que me ayudaron a cumplir una de mis metas.

A todos ellos, gracias.

Resumen

En este trabajo de tesis el filtro de Kalman como un algoritmo que permite realizar el acondicionamiento de la señal proporcionada por un instrumento, que por su naturaleza es fácilmente contaminado por ruido, este instrumento es un acelerómetro, el cual en combinación con un giroscopio forman un sistema dinámico cuyo modelo matemático es de tipo lineal, estas características lo hacen ideal para aplicar el algoritmo del filtro de Kalman. Dado que el filtro de Kalman es un estimador de estados de varianza mínima, se incluye un capítulo donde se presenta una recopilación de conceptos básicos de la teoría de la probabilidad que son necesarios para entender mejor el funcionamiento del filtro de Kalman. Además de la aplicación a la medición de inclinación mediante acelerómetro y giroscopio, la cual se valida en simulación, también se presenta un caso de estudio ilustrativo consistente en la aplicación del filtro de Kalman a la estimación óptima de estados en un sistema de nivel de líquido.

Índice general

Agradecimientos	I
Resumen	III
Índice de figuras	IX
Índice de cuadros	XI
Nomenclatura	XIII
Capítulo 1. Introducción	1
1.1. Antecedentes	1
1.2. Descripción general del problema	1
1.3. Objetivos	2
1.4. Justificación	2
1.5. Metodología	2
1.6. Estructura de la tesis	3
Capítulo 2. Bases de la teoría de la probabilidad	5
2.1. Experimentos aleatorios	5
2.2. Espacios muestrales	6
2.3. Eventos	6
2.4. El concepto de probabilidad	7
2.5. Los axiomas de probabilidad	8
2.6. Teoremas básicos en probabilidad	9
2.7. Probabilidad condicional	10
2.8. Teorema de Bayes	11
2.9. Análisis Combinatorio	12
2.10. Permutaciones	12
2.11. Combinaciones	13
2.12. Variables aleatorias	14
2.13. Distribuciones de probabilidad discretas	15

2.14.	Funciones de distribución acumulativa	16
2.15.	Funciones de distribución para variables aleatorias discretas	16
2.16.	Variables aleatorias continuas	18
2.17.	Obtención de $p(x)$ a partir de $P(X)$	21
2.18.	Interpretación gráfica	21
2.19.	Distribuciones de probabilidad conjunta	22
2.20.	Caso de dos variables aleatorias discretas	23
2.21.	Variables aleatorias independientes	25
2.22.	Distribuciones condicionales	26
2.23.	Esperanza matemática	27
2.24.	Esperanza de funciones de variables aleatorias	29
2.25.	Teoremas de la esperanza matemática	30
2.26.	La varianza y la desviación estandar	30
2.27.	Teoremas de la varianza	31
2.28.	Variables aleatorias estandarizadas	32
2.29.	Momentos	32
2.30.	Función generadora de momentos	33
2.31.	Media, varianza y covarianza para distribuciones conjuntas	34
2.32.	Teoremas de la covarianza	34
2.33.	Coefficiente de correlación	35
2.34.	Otras medidas de tendencia central	35
2.35.	Desviación media	35
2.36.	Sesgo	36
2.37.	Kurtosis	37
2.38.	Distribución binomial	37
2.39.	Distribución normal	38
2.40.	Procesos estocásticos	40
2.41.	Ruido blanco gaussiano	41
Capítulo 3.	Filtro de Kalman	43
3.1.	Filtro de kalman discreto	43
3.2.	El proceso de la estimación	43
3.3.	La estrategia computacional del filtro de Kalman	44
3.4.	Las propiedades probabilísticas del filtro de Kalman	46
3.5.	El algoritmo discreto del filtro de Kalman	46
3.6.	El filtro de Kalman como observador de estados	48

3.7. Ejemplo Sistema de nivel del líquido	50
Capítulo 4. Aplicación del Filtro de Kalman a la medición de inclinación	57
4.1. Acelerómetros	57
4.2. Giroscopio	61
4.3. Aplicación del filtro de kalman	63
4.4. Conclusiones	67
Capítulo 5. Conclusiones	69
5.1. Conclusiones Generales	69
Bibliografía	71

Índice de figuras

2.1	Diagramas Venn para ilustrar la intersección y la unión de dos eventos A,B	10
2.2	Función de probabilidad discreta	16
2.3	Gráfica de una función de distribución acumulativa (para el Ejemplo 16.1)	17
2.4	Función de densidad de probabilidad del ejemplo 16.2.	20
2.5	Función de densidad de probabilidad del ejemplo 16.3.	21
2.6	Se ilustra la función de densidad de probabilidad (pdf)	22
2.7	Se ilustra la función de distribución acumulativa	22
2.8	Se ilustra la superficie de probabilidad y la probabilidad $\Pr(a < X < b, c < Y < d)$ 24	31
2.9	Ilustrando diferentes varianzas y la misma media.	36
2.10	Sesgado a la derecha	36
2.11	Sesgado a la izquierda	36
2.12	Kurtosis	37
2.13	Curva normal estándar.	40
3.1	Etapas de predicción y corrección del filtro de Kalman	47
3.2	Imagen completa de la operación del filtro de Kalman.	48
3.3	Observador de estados tipo Luenberger	49
3.4	El filtro de Kalman como un observador de Luenberger más optimizador de la ganancia K_k	50
3.5	Sistema de nivel de líquido	51
3.6	Diagrama a bloques del sistema de nivel de líquido	52
3.7	Resultados de la simulación	56
4.1	Resorte y Bola	57
4.2	Sismógrafo	58

4.3	Acelerómetro electrónico	58
4.4	Acelerómetro (Un sólo eje) para medir inclinación	59
4.5	Acelerómetro (Dos ejes) para medir inclinación.	59
4.6	Descomposición vectorial.	60
4.7	Gráfica de la aproximación del seno	61
4.8	Ejemplo de un giroscopio	62
4.9	Rueda girando sobre un eje mostrando las fuerzas de reacción por efecto giroscopio.	62
4.10	Resultado de la simulación.	67

Índice de cuadros

2.1	Muestras Asociadas a X	14
2.2	Tipos de procesos estocásticos (Continuos y discretos).	41

Nomenclatura

pdf	Función de distribución de probabilidad (en el caso discreto) y función de densidad de probabilidad (en el caso continuo)
S	Espacio Muestral
FK	Filtro de Kalman

Capítulo 1

Introducción

1.1. Antecedentes

De las herramientas matemáticas que pueden usarse para estimación estocástica de mediciones ruidosas de sensores, una de las más conocidas y de las más usadas es la que se conoce como el Filtro de Kalman. El filtro de Kalman se llama así por Rudolph E. Kalman, quien en 1960 publicó su famoso artículo describiendo una solución recursiva para el filtrado lineal discreto Kalman (1960). Una muy amigable introducción a la idea general del filtro de Kalman se ofrece en el capítulo 1 del libro de Peter (1979). Referencias más amplias pueden ser encontradas en Ham and Brown (1983); Jacobs and Jacobs (1993).

El filtro de Kalman es esencialmente un conjunto de ecuaciones matemáticas que implementan un estimador del tipo predictor-corrector, basado en un modelo de espacio lineal, el cuál es óptimo en el sentido de que minimiza la varianza del error de estimación, cuando se presenta un conjunto de condiciones adecuadas que se detallan más adelante. Se han hecho investigaciones extensas en el área de navegación autónoma o asistida. Pero fueron los avances en la computación digital, los que hicieron capaces que el filtro fuera práctico, pero también tuvo que ver su naturaleza simple y robusta (se resiste a las perturbaciones externas). Raramente todas las condiciones necesarias para hacerlo óptimo se consiguen y aún así, el filtro funciona aparentemente bien para muchas aplicaciones.

1.2. Descripción general del problema

Enfocaremos esta tesis a la solución de un problema muy común en el área de la instrumentación electrónica, éste se refiere a la contaminación que alguna señal puede adquirir durante el proceso de adquisición, transmisión o recepción. Se busca comprobar el funcionamiento del filtro de Kalman para este tipo de problemas, especialmente se tiene interés en su aplicación al filtrado de las señales producidas por un acelerómetro y el uso de un giroscopio para mejorar la medición.

1.3. Objetivos

El principal objetivo de este trabajo es probar en simulación el desempeño del filtro de Kalman utilizado como un filtro del ruido en la medición de inclinación angular realizada mediante un acelerómetro-giroscopio electrónicos. Para lograr este objetivo principal es necesario alcanzar primeramente los siguientes objetivos secundarios:

- Revisar las bases de la teoría de la probabilidad que son utilizadas en la deducción del filtro de Kalman.
- Elaborar un resumen, de estas bases teóricas, comprensible para un estudiante de ingeniería.
- Presentar el algoritmo de estimación del filtrado de Kalman de una manera comprensible.
- Evaluar en simulación el algoritmo para algunos ejemplos teóricos.
- Evaluar en simulación del algoritmo para la medición de inclinación angular basada en un acelerómetro y giroscopio electrónicos.

1.4. Justificación

El filtro de Kalman es un algoritmo óptimo muy utilizado en todo lo que se refiere al procesamiento de señales que involucran variables aleatorias. Un ejemplo claro es el de los sistemas de navegación que se utiliza en los automóviles nuevos, tales son por ejemplo el Sistema de Posicionamiento Global (GPS), así como un Sistema de Navegación Inercial (INS) para ayudar al conductor a encontrar el camino.

Por otro lado, el uso de acelerómetro y giroscopio se ha generalizado en los dispositivos electrónicos actuales, tales como tabletas, electrónicas y teléfonos celulares, por esta razón es necesario implementar algoritmos efectivos de filtrado, tales como el filtro de Kalman.

1.5. Metodología

Con el fin de comprender de una manera un poco mas profunda el filtro de Kalman;

- se analizarán algunas de sus bases probabilísticas,
- se entenderá el porqué de las ecuaciones que lo describen y
- finalmente se hará una aplicación a la medición de inclinación, utilizando señales recibidas de un acelerómetro y un giroscopio.

1.6. Estructura de la tesis

La tesis está seccionada en 4 capítulos de la siguiente forma:

- Capítulo 1 Introducción: Habla de los antecedentes, la descripción general del problema, los objetivos, la justificación, la metodología y la estructura de la tesis.
- Capítulo 2 Bases de la teoría de la probabilidad: Menciona los conceptos probabilísticos principales para entender en que se basa el filtro de Kalman.
- Capítulo 3 Filtro de Kalman: Se describe el funcionamiento del filtro de Kalman, así como las ecuaciones que lo describen y se presenta un ejemplo típico de aplicación del filtro de Kalman como un observador de estados de varianza mínima en presencia de ruido.
- Capítulo 4 Aplicación del filtro de Kalman a la medición de inclinación: Se presenta un preámbulo de los elementos para medir inclinación y una simulación de un sistema de medición de inclinación filtrado por el filtro de Kalman.
- Capítulo 5 Conclusiones generales de la tesis.

Bases de la teoría de la probabilidad

Ya que el Filtro de Kalman es un algoritmo óptimo para procesar datos, tiene relación con muchas disciplinas en el área de las matemáticas, una de ellas (quizá con la que tiene mayor relación) es la de la probabilidad. De ahí el hecho que se dedique un capítulo completo a las bases de la teoría de la probabilidad necesaria para comprender de mejor manera al Filtro de Kalman. La mayor parte de la teoría y sus ejemplos presentados en este capítulo han sido tomados de Papoulis and Pillai (2002)

2.1. Experimentos aleatorios

Es ampliamente reconocida la enorme importancia de los experimentos en la ciencia e ingeniería. Sin embargo, en algunos experimentos, no es posible asegurar o controlar el valor de algunas variables, así que el resultado variará de una ejecución del experimento, con respecto a otra, aún cuando las condiciones sean las mismas. Estos experimentos son descritos como aleatorios.

EJEMPLO. 1.1 Si lanzamos un moneda, el resultado del experimento tiene dos opciones, que salga “Cruz”, que se representará por una T (ó por un 0), o que salga una “Cara”, que se representará por una H (ó por un 1). Así los elementos del conjunto de resultados serán $\{H,T\}$ ó $\{0,1\}$.

EJEMPLO. 1.2 Si lanzamos un dado, lo obtenido en el experimento puede ser cualquiera de los números del conjunto de resultados $\{1,2,3,4,5,6\}$.

EJEMPLO. 1.3 Si lanzamos una moneda dos veces, hay cuatro posibles resultados, como se indica en el siguiente conjunto de resultados posibles $\{HH,HT,TH,TT\}$.

En el ejemplo 2.3 se observa como el resultado de un experimento aleatorio, repetido dos veces en éste caso, hace que nuestro conjunto de posibles resultados se haga mas grande (Nótese que el orden en el que se obtuvo el resultado sí importa).

2.2. Espacios muestrales

Un conjunto S que consiste en todos los posibles resultados de un experimento aleatorio es llamado espacio muestral, y cada uno de los resultados es llamado muestra (en el espacio muestral S). A menudo habrá mas de un espacio muestral que puede describir resultados de un experimento, pero casi siempre habrá uno que proveerá la mayor cantidad de información.

EJEMPLO. 2.1 Si lanzamos un dado, un espacio muestral, o conjunto de resultados, está dado por $\{1,2,3,4,5,6\}$, mientras que otro sería $\{\text{Non,Par}\}$

Es usualmente útil representar un espacio muestral gráficamente. En esos casos, es recomendable usar números en lugar de letras cuando sea posible.

EJEMPLO. 2.2 Si lanzamos un moneda dos veces y se utiliza la misma nomenclatura que en el ejemplo 2.1. El espacio muestral puede ser representado por los siguientes puntos:

$\{(0,1) (1,1),(0,0) (1,0)\}$, o bien $\{(T,H) ,(H,H),(T,T) ,(H,T)\}$

DEFINICIÓN. 2.1 Espacio muestral finito: Tiene un numero finito de muestras como en el ejemplo 2.5

DEFINICIÓN. 2.2 Espacio muestral infinito contable: tiene tantas muestras como numeros naturales $\{1,2,3,\dots\}$, por ejemplo el número de veces que se requiere lanzar un dado hasta obtener un número 6.

DEFINICIÓN. 2.3 Espacio muestral infinito no contable: Tiene tantas muestras como hay en algún intervalo del eje de las X (Continuo). Por ejemplo la coordenada horizontal x del impacto de un tiro al blanco $x_{min} \leq x \leq x_{max}$

2.3. Eventos

Un evento es un subconjunto A del espacio muestral S . A es un conjunto de posibles resultados. Si el resultado del experimento es un elemento de A , decimos que el evento A ha ocurrido. Un evento que consiste solamente de una muestra de S , es llamada un evento simple o un evento elemental.

EJEMPLO. 3.1 Si lanzamos una moneda dos veces, el evento de que solo se obtenga una cara, es el subconjunto del espacio muestral que consiste en las muestras $(0,1)$ y $(1,0)$ del ejemplo 2.5.

DEFINICIÓN. **3.1 Evento seguro:** Es el mismo espacio muestral S , ya que un elemento de S siempre debe ocurrir.

DEFINICIÓN. **3.2 Evento imposible:** También llamado el conjunto vacío \emptyset porque un elemento de \emptyset no puede ocurrir.

Usando las operaciones de eventos en S , podemos obtener otros eventos en S . Por ejemplo, si A y B son eventos (subconjuntos de S), entonces la suma lógica, el producto lógico y el complemento lógico produce los siguientes nuevos eventos:

1. **Suma lógica (OR lógico):** Es la unión de A y B , por definición, la unión $A \cup B$ es el evento “cualquiera A o B o los dos”.
2. **Producto lógico (AND lógico):** Es la intersección de A y B , por definición, la intersección $A \cap B$ es el evento “ambos A y B ”.
3. **Complemento lógico (NOT lógico):** Es el complemento de A , por definición, el complemento \bar{A} es el evento “no es A ”.
4. **Diferencia lógica:** Por definición, $A - B = A \cap \bar{B}$ es el evento “ A pero no B ”, en particular, $\bar{A} = S - A$.
5. **Eventos mutuamente excluyentes:** Son dos eventos A, B de los cuales sus conjuntos correspondientes son disjuntos en S . $A \cap B = \emptyset$. Esto quiere decir que no pueden ocurrir los dos al mismo tiempo.

EJEMPLO. **3.2** Refiriéndose al experimento de lanzar la moneda dos veces, sea A el evento “que al menos una cara ocurra” y B el evento “que el segundo lanzamiento sea una cruz”. Se definen dos eventos

$$A = HT, TH, HH \text{ y } B = HT, TT$$

Entonces aplicando las operaciones lógicas previamente definidas, tenemos:

$$A \cup B = HT, TH, HH, TT = S$$

$$A \cap B = HT$$

$$\bar{A} = TT$$

$$A - B = A \cap \bar{B} = TH, HH$$

2.4. El concepto de probabilidad

En cualquier experimento aleatorio, siempre hay cierto grado de incertidumbre de si un evento en particular ocurrirá o no. Para medir la oportunidad, o probabilidad, con la que se puede esperar que el evento ocurra, es conveniente asignarle un número entre 0 y 1, o bien entre el 0 % y el 100 %.

Si se sabe que el evento siempre ocurrirá, decimos que su probabilidad es 100 % o 1.

Si se sabe que el evento no ocurrirá, decimos que su probabilidad es 0 % o 0.

Si por ejemplo la probabilidad es $\frac{1}{4}$, se está seguro de que hay un 25 % de probabilidad de que ocurra y un 75 % de probabilidad de que no ocurra. De igual forma, se dice que las probabilidades de que no ocurra contra las que sí ocurra son de 75 % a 25 %, o bien de 3 a 1.

Hay dos enfoques para definir la probabilidad de un evento.

1. **Enfoque clásico (axiomático):** Si un evento puede ocurrir de n diferentes maneras, de un total de N posibles que pueden ocurrir en el experimento, donde todas tienen la misma probabilidad de ocurrir, entonces la probabilidad del evento se puede expresar como $\frac{n}{N}$.
2. **Enfoque por frecuencia.** Si después de N repeticiones de un experimento, donde N es un número muy grande, un evento se observa que ocurre en n de estas repeticiones del experimento, entonces la probabilidad de que el evento ocurra es $\frac{n}{N}$.

EJEMPLO. 4.1 En el experimento de lanzar una moneda, hay dos posibles formas de que la moneda caiga {T o H}, podemos razonar que la probabilidad requerida es $\frac{1}{2}$ (asumimos que la moneda no está cargada). Este tipo de cálculo de probabilidad, es un ejemplo de la aproximación clásica.

EJEMPLO. 4.2 Si lanzamos una moneda 1000 veces y vemos que la cara de la moneda sale 520 veces, podemos estimar la probabilidad del evento H (que salga cara) sea de $\frac{520}{1000} = \frac{52}{100}$. Este tipo de cálculo de probabilidad, es un ejemplo de la aproximación por frecuencia.

Ambas aproximaciones tienen desventajas fundamentales, la primera, porque las palabras “tienen la misma probabilidad de ocurrir” son vagas y la segunda porque el “número muy grande” involucra algo vago también.

Es por eso que los matemáticos han llegado a una aproximación axiomática de la probabilidad.

2.5. Los axiomas de probabilidad

Suponiendo que tenemos un espacio muestral S . Si S es discreto, todos los subconjuntos corresponden a eventos, pero si S es no discreta, solo subconjuntos especiales (llamados medibles) corresponden a eventos. Para cada evento A en la

clase C de eventos, se puede asociar, el número real $\Pr(A)$, donde \Pr es la función de probabilidad y $\Pr(A)$ la probabilidad del evento A , si los siguientes axiomas se satisfacen:

AXIOMA. **1** Para cada evento A en la clase C ,

$$\Pr(A) \geq 0 \quad (2.1)$$

AXIOMA. **2** Para el evento seguro S in la clase S ,

$$\Pr(S) = 1 \quad (2.2)$$

AXIOMA. **3** Para cualquier número de eventos mutuamente excluyentes A_1, A_2, \dots in la clase C ,

$$\Pr(A_1 \cup A_2 \cup \dots) = \Pr(A_1) + \Pr(A_2) + \dots \quad (2.3)$$

2.6. Teoremas básicos en probabilidad

TEOREMA. **1** Para cualquier evento A

$$0 \leq \Pr(A) \leq 1 \quad (2.4)$$

La probabilidad se encuentra entre 0 y 1

TEOREMA. **2**

$$\Pr(\emptyset) = 0 \quad (2.5)$$

El evento imposible tiene probabilidad cero

TEOREMA. **3** Si \bar{A} es el complemento de A , entonces

$$\Pr(\bar{A}) = 1 - \Pr(A) \quad (2.6)$$

TEOREMA. **4** Si $A = A_1 \cup A_2 \cup \dots \cup A_n$ donde A_1, A_2, \dots, A_n son elementos mutuamente excluyentes entonces

$$\Pr(A) = \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_n) \quad (2.7)$$

TEOREMA. **5** si A y B son cualquier par de eventos, entonces

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \quad (2.8)$$

TEOREMA. 6 Para dos eventos cualesquiera A y B

$$Pr(A) = Pr(A \cap B) + Pr(A \cap \bar{B}) \quad (2.9)$$

Todos estos teoremas pueden ser fácilmente probados usando un diagrama de Venn, como se ve en la figura 2.1, para ilustrar las operaciones unión e intersección de eventos:

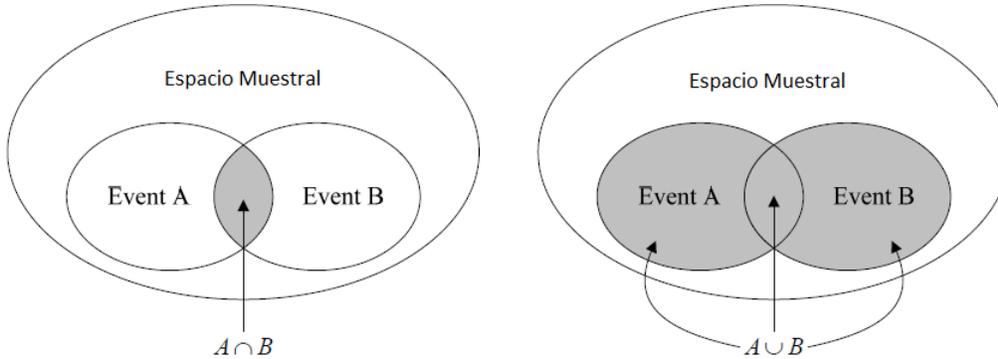


FIGURA 2.1. Diagramas Venn para ilustrar la intersección y la unión de dos eventos A,B

2.7. Probabilidad condicional

Sean A y B dos eventos tales que $Pr(A) > 0$. Denotamos por $Pr(B|A)$ la probabilidad de B dado que el evento A ya ocurrió. Como A se sabe que ya ocurrió, se convierte en el nuevo espacio muestral reemplazando al original S. En este nuevo espacio muestral, por definición (la tan llamada regla de Bayes)

$$Pr(A \cap B) = Pr(A)Pr(B|A) \quad (2.10)$$

Esto da como resultado la definición de la probabilidad condicional

$$Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)} \quad (2.11)$$

En palabras, 2.10 dice que la probabilidad de que ambos A y B ocurran es igual a la probabilidad de que A ocurra, por la probabilidad de que B ocurra dado que A ya ocurrió. Podemos llamar $Pr(B|A)$ la probabilidad condicional de B dado A. Y por supuesto, la probabilidad condicional también satisface los axiomas de probabilidad.

EJEMPLO. 7.1 Encuentra la probabilidad, de que en un solo lanzamiento de un dado resulte un número menor que 4, si se sabe que el resultado del lanzamiento es un número impar.

El evento que ocurrió A, es un número impar $\{1 \text{ o } 3 \text{ o } 5\}$. Así para probabilidades iguales

$$Pr(A) = Pr(1 \cup 3 \cup 5) = 1/6 + 1/6 + 1/6 = 1/2$$

El evento B está formado por los resultados menores que 4, es decir, $B = \{1, 2, 3\}$ por lo tanto $A \cap B = \{1, 3\}$

$$\text{Entonces, la probabilidad } Pr(A \cap B) = 2/6 = 1/3$$

Por lo tanto, la probabilidad de $1 \cup 3$ que sea menor que 4, dado que el resultado del lanzamiento es un número impar. Así

$$Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Por tanto, dado el conocimiento que teníamos (información a priori) de que el resultado del lanzamiento era un número impar, la probabilidad aumenta de $P(B) = 1/2$ a $Pr(B|A) = 2/3$

Si A y B son mutuamente excluyentes, entonces $Pr(B|A) = 0$

Por otro lado, si A está contenido dentro de B, entonces $Pr(B|A) = 1$

2.8. Teorema de Bayes

Suponiendo que A_1, A_2, \dots, A_J son eventos mutuamente excluyentes cuya unión es el espacio muestral, la suposición nos dice que uno de los eventos debe ocurrir. Entonces si A es cualquier evento, el siguiente teorema es importante (La regla de Bayes) es válida

$$Pr(A_k|A) = \frac{Pr(A_k)Pr(A|A_k)}{\sum_{j=1}^J Pr(A_j)Pr(A|A_j)} \quad (2.12)$$

Esto nos permite encontrar la probabilidad de que alguno de varios eventos A_1, A_2, \dots, A_J que pueden ocasionar que A ocurra. Sea el que lo haya provocado, por esta razón el teorema de Bayes, se llama también el teorema de probabilidad de las causas.

EJEMPLO. 8.1 Un doctor dispone de tres equipos electrónicos para realizar ecosonogramas. El uso que le da a cada equipo es de 25% al primero, 35% al segundo y 40% al tercero. Se sabe que los aparatos tienen probabilidades de error de 1%, 2% y 3% respectivamente. Un paciente busca el resultado de un ecografía

y observa que tiene un error. Determine la probabilidad de que se ha usado el primer aparato.

Se definen los eventos.

Evento P: Seleccionar el primer aparato.

Evento S: Seleccionar el segundo aparato.

Evento T: Seleccionar el tercer aparato.

Evento E: Seleccionar un resultado con error.

Se observa que se trata de determinar la probabilidad de que un examen errado sea del primer aparato, es decir, ya ha ocurrido el error. Por lo tanto, debemos recurrir al teorema de Bayes. Entonces se tiene:

$$P(P | E) = \frac{P(P) * P(E | P)}{P(P) * P(E | P) + P(S) * P(E | S) + P(T) * P(E | T)} \quad (2.13)$$

$$P(P | E) = \frac{0.25 * 0.01}{0.25 * 0.01 + 0.35 * 0.02 + 0.4 * 0.03} = 0.116 = 11.6\% \quad (2.14)$$

2.9. Análisis Combinatorio

En muchos casos el número de muestras en el espacio muestral no es muy grande, así que directamente de enumerar o contar las muestras, se pueden obtener las probabilidades sin mayor dificultad.

Sin embargo, los problemas se agrandan cuando contar directamente se vuelve prácticamente imposible. En esos casos, se requiere un análisis combinatorio que ofrece fórmulas para contar las llamadas permutaciones, combinaciones y otros casos similares.

2.10. Permutaciones

Dados N distintos objetos, queremos acomodar r de ellos en línea. Como hay N formas de escoger el primer objeto, y después, N-1 formas de escoger el segundo, y finalmente N-r+1 formas de escoger el r-ésimo objeto, siguiendo el principio fundamental de conteo, de acuerdo al cual el número de diferentes acomodados está dado por el producto del número de formas de escoger cada uno de ellos, entonces

$${}_N P_r = N(N - 1) \dots (N - r + 1) \quad (2.15)$$

En el caso particular donde r=N, tenemos

$${}_N P_r = N!$$

que es llamado N-factorial. Así en términos de factoriales, 2.15 se convierte en

$${}_N P_r = \frac{N!}{(N-r)!} \quad (2.16)$$

Nótese que en las permutaciones estamos interesados en el orden de los objetos. Por ejemplo ABC es un acomodo diferente que BCA y diferente que CAB y diferente de CBA, etc.

EJEMPLO. 10.1 Se dará un premio entre las mejores 2 personas de 10 participantes. ¿De cuantas diferentes maneras se puede dar el primero y segundo lugar?

$${}_{10}P_2 = \frac{10!}{(10-2)!} = \frac{10 * 9 * 8!}{8!} = 90 \quad (2.17)$$

2.11. Combinaciones

En muchos problemas, estamos interesados solo en seleccionar objetos sin importar el orden. A este tipo de selecciones, se les llama combinación. Así, por definición:

DEFINICIÓN. 11.1 Combinaciones: son acomodos de varios objetos sin importar el orden.

Por ejemplo, ABC, BCA, CBA, CAB, son la misma combinación.

El numero total de combinaciones de r objetos seleccionados de n objetos en total (las combinaciones de n cosas tomadas r a la vez)

$${}_N C_r = \binom{N}{r} = \frac{N!}{(N-r)!r!} = \frac{N(N-1)\dots(N-r+1)}{r!} \quad (2.18)$$

También puede ser expresada como:

$${}_N C_r = \frac{{}_N P_r}{r!} \quad (2.19)$$

Por cálculos directos, puede mostrarse que

$$\binom{N}{r} = {}_N C_r = {}_N C_{N-r} = \binom{N}{N-r} \quad (2.20)$$

EJEMPLO. 11.1 El número de maneras en la que 3 cartas pueden ser escogidas de un total de 8 diferentes cartas, es

$${}_8C_3 = \binom{8}{3} = \frac{8*7*6}{3!} = \frac{8*7*6}{1*2*3} = \frac{8*7*6}{6} = 56$$

2.12. Variables aleatorias

DEFINICIÓN. 12.1 Variable aleatoria: Es una función de valor real definida en el espacio muestral.

Una variable aleatoria puede asociarse con una muestra en algún espacio muestral S . Usualmente podemos suponer que cada punto del espacio muestral S le asignamos un número. Una variable aleatoria se denota como una letra mayúscula, como X o Y .

En general, una variable aleatoria que suele tener alguna especificación física, geométrica, o de algún otro tipo de acuerdo al problema de interés.

EJEMPLO. 12.1 Suponiendo que una moneda es lanzada dos veces, para que el espacio muestral sea

$$S = \{HH, HT, TH, TT\}$$

Sea X , la variable aleatoria, que representa el número de caras que pueden salir. Con cada muestra podemos asociar un número para X como se muestra en el Cuadro 2.1 de abajo:

Muestras	HH	HT	TH	TT
X=Numero de caras	2	1	1	0

CUADRO 2.1. Muestras Asociadas a X

Por ejemplo, en el caso de que el evento HH , $X=2$, mientras que para el evento TH , $X=1$. Así, debido a que todos los eventos HH, HT, TH, TT , son aleatorios, el valor que tome X es aleatorio.

Debe notarse que diferentes variables aleatorias pueden ser definidas en el mismo espacio muestral $S = \{HH, HT, TH, TT\}$, por ejemplo, el cuadrado de número de caras, o el número de caras menos el número de cruces, etc.

DEFINICIÓN. 12.2 Variable aleatoria discreta: Es una variable aleatoria que toma un número finito o un número infinito contable de valores.

DEFINICIÓN. 12.3 Variables aleatoria no discreta (continua): Es una variable que toma un número infinito no contable de valores (en algún intervalo continuo).

2.13. Distribuciones de probabilidad discretas

Sea X un variable aleatoria discreta. Suponiendo que los posibles valores que puede tomar están dados por x_1, x_2, x_3, \dots , acomodados en algún orden. Entonces, podemos asociar estos valores con probabilidades dadas por

$$p(x_k) \geq 0 = Pr(X = x_k); k = 1, 2, 3, \dots \quad (2.21)$$

DEFINICIÓN. 13.1 Distribución de probabilidad: Esta definida como un función de probabilidad para cada posible valor x de la variable aleatoria X , como sigue

$$p(x) = Pr(X = x) \quad (2.22)$$

Si x_k está en el espacio muestral, $p(x) = p(x_k) = Pr(X = x_k)$; para otros valores de x que no están en el espacio muestral, $p(x) = 0$

En general, $p(x)$ es una función de probabilidad si

1. $p(x) \geq 0$
2. $\sum_x p(x) = 1$, donde la suma es tomada de todos los posibles valores de x .

EJEMPLO. 13.1 La función de probabilidad correspondiente a la variable aleatoria X del Ejemplo 12.1. Asumiendo que la moneda no está cargada, se obtiene lo siguiente

$$Pr(HH) = 1/4, Pr(HT) = 1/4, Pr(TH) = 1/4, Pr(TT) = 1/4$$

Entonces para la variable aleatoria X se tiene que

$$p(0) = Pr(X = 0) = Pr(TT) = 1/4$$

$$p(1) = Pr(X = 1) = Pr(HT \cup TH) = Pr(HT) + Pr(TH) = 1/4 + 1/4 = 1/2$$

$$p(2) = Pr(X = 2) = Pr(HH) = 1/4$$

La función de probabilidad de este ejemplo se puede representar en forma gráfica como se muestra en la figura 2.2.

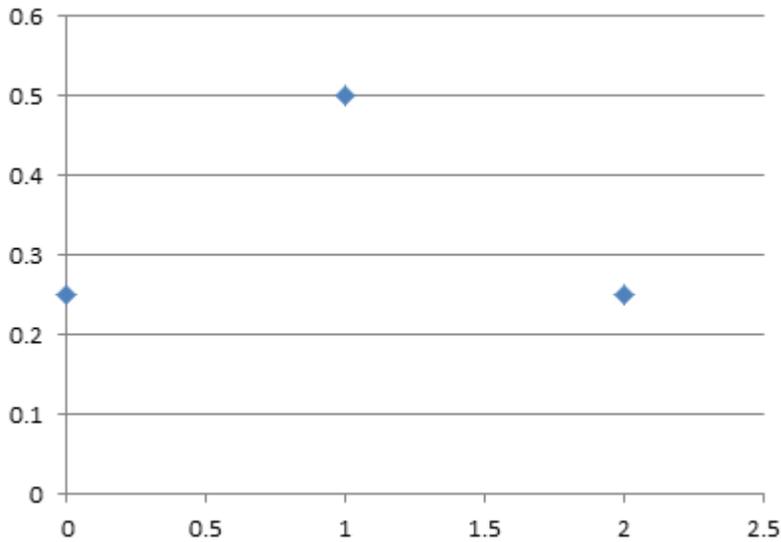


FIGURA 2.2. Función de probabilidad discreta

2.14. Funciones de distribución acumulativa

DEFINICIÓN. 14.1 Función de distribución acumulativa (cdf): Para cualquier variable aleatoria X , en el caso general, esta definida por

$$P(x) = Pr(X \leq x), \text{ donde } x \text{ es cualquier número real } -\infty < x < \infty \quad (2.23)$$

La función de distribución acumulativa $P(x)$, tiene las siguientes propiedades:

1. $P(x) \geq 0$ es no decreciente, es decir, $P(x) \leq P(y)$ si $x \leq y$
2. $\lim_{x \rightarrow -\infty} P(x) = 0$; $\lim_{x \rightarrow +\infty} P(x) = 1$
3. $P(x)$ es continua desde la derecha, $\lim_{h \rightarrow 0^+} P(x+h) = P(x)$ para toda x .

2.15. Funciones de distribución para variables aleatorias discretas

La función de distribución acumulativa para un variable aleatoria X , puede ser obtenida de su función de probabilidad notando que, para toda x en $(-\infty, \infty)$

$$P(x) = Pr(X \leq x) = \sum_{u \leq x} p(u) \quad (2.24)$$

Donde la suma es tomada sobre todos los valores u , para cuando $u \leq x$.

Si X toma solo un número finito (N) de valores $x_1, x_2, \dots, x_n, \dots, x_N$, entonces la función de distribución acumulativa está dada por:

$$P(x) = \begin{cases} P(x) = 0 & -\infty < x < x_1 \\ P(x) = p(x_1) & x_1 \leq x < x_2 \\ P(x) = p(x_1) + p(x_2) & x_2 \leq x < x_3 \\ \dots & \dots \\ P(x) = p(x_1) + p(x_2) + \dots + p(x_N) & x_N \leq x < \infty \end{cases} \quad (2.25)$$

EJEMPLO. 15.1

- (a) Encontrar la función de distribución acumulativa para la variable aleatoria X del Ejemplo 13.1.
- (b) Obtener su grafica.

Solución para (a): Usando los datos del ejemplo 12.1 y aplicando la ecuación 2.25, se obtiene la función de distribución para la variable aleatoria X:

$$P(x) = \begin{cases} P(x) = 0 & -\infty < x < 1 \\ P(x) = 1/4 & 0 \leq x < 1 \\ P(x) = 3/4 & 1 \leq x < 2 \\ P(x) = 1 & 2 \leq x < \infty \end{cases}$$

Solución para (b): La gráfica de P(x) se muestra en la Figura 2.3

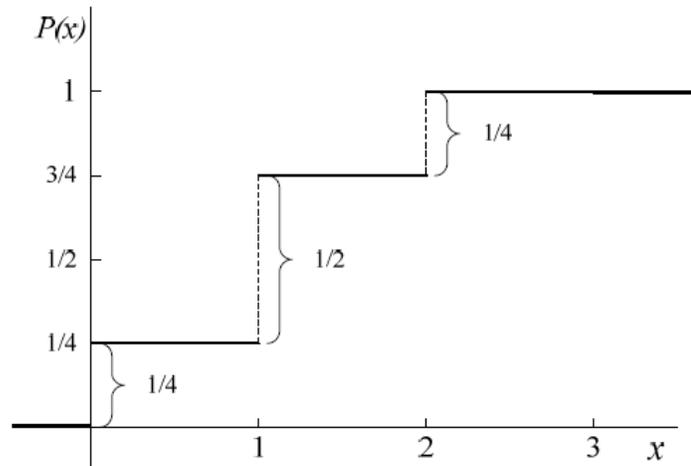


FIGURA 2.3. Gráfica de una función de distribución acumulativa (para el Ejemplo 16.1)

Las siguientes observaciones acerca de la función de distribución de arriba, las cuales son ciertas en general, deben notarse.

1. Las magnitudes de los saltos en 0,1,2 son $1/4$, $1/2$, $1/4$, son precisamente las probabilidades de la figura 2.2. Este hecho, nos permite obtener la función de probabilidad acumulativa a partir de la función de distribución.
2. Debido a la apariencia de la gráfica en la Figura 2.3, usualmente es llamada función de escalera o escalonada. El valor de la función en un entero es obtenida del paso mas alto, asi el valor de 1 es $3/4$, y no $1/4$, Esto es expresado matemáticamente diciendo que la función de distribución es continua desde la derecha en 0, 1, 2.
3. Mientras nos acercamos de la izquierda a la derecha (“subiendo la escalera”), la función de distribución puede mantenerse o incrementar, tomando valores desde 0 hasta 1. Debido a esto, se dice que es una función monótonicamente creciente.

La función de probabilidad de un variable aleatoria discreta puede ser obtenida de su función de distribución acumulativa, tomando en cuenta que

$$p(x) = P(x) - \lim_{u \rightarrow x^-} P(u) \quad (2.26)$$

2.16. Variables aleatorias continuas

DEFINICIÓN. 16.1 Variable aleatoria continua: Es una variable aleatoria que puede tomar un conjunto de valores infinito y no numerable cuya función de distribución acumulativa puede ser representada como

$$P(x) = Pr(X \leq x) = \int_{-\infty}^x p(u)du, \text{ para } -\infty < x < \infty \quad (2.27)$$

Donde la función $P(x)$ tiene las siguientes propiedades:

1. $P(x) \geq 0$ es no decreciente, por tanto, $P(x) \leq P(y)$ si $x \leq y$
2. $\int_{-\infty}^{\infty} p(x)dx = 1$

Por tanto, si X es una variable aleatoria continua, entonces la probabilidad de que X tome cualquier valor particular es cero, mientras que la probabilidad de que X tome un valor dentro de en un intervalo, es decir, que se coloque entre dos diferentes valores, digamos, a y b , esta dada por

$$Pr(a < X < b) = \int_a^b p(x)dx \quad (2.28)$$

EJEMPLO. 16.1 Si un individuo es seleccionado al azar de un gran grupo de adultos, hombres, la probabilidad de que su altura X sea exactamente 175.452168cm sería cero. Sin embargo, hay una probabilidad mayor a cero de que X esté entre 170 cm y 180 cm, por ejemplo.

DEFINICIÓN. 16.2 Función de densidad de probabilidad (o pdf): es una función $p(x)$ que satisface 2.272.28, los requerimientos de las propiedades para una variable continua aleatoria.

Cualquier función $p(x)$ definida por 2.27 que satisface las propiedades 1 y 2 de arriba automáticamente será una función de densidad de probabilidad, las probabilidades requeridas podrán ser obtenidas de 2.28.

EJEMPLO. 16.2

(a) Encuentre la constante c , tal que la función:

$$p(x) = \begin{cases} cx^2 & 0 \leq x < 3 \\ 0 & \text{En otro caso} \end{cases}$$

es una función de densidad de probabilidad y graficarla.

(b) Calcular la probabilidad $Pr(1 < X < 2)$

Solución para (a): Como $p(x)$ debe satisfacer la propiedad 1 entonces $c \geq 0$, por otro lado, como debe satisfacer la propiedad 2, entonces

$$\int_{-\infty}^{\infty} p(x)dx = \int_0^3 cx^2dx = \frac{cx^3}{3} \Big|_0^3 = 9c = 1$$

Por lo tanto $c = 1/9$

Solución para (b):

$$Pr(1 < X < 2) = \int_1^2 \frac{1}{9}x^2dx = \frac{x^3}{27} \Big|_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}$$

En caso de que $p(x)$ sea continuo, lo cual debemos asumir a menos que otra cosa sea dicha, la probabilidad de que X sea igual a cualquier valor particular es cero. En tal caso podemos reemplazar ambos símbolos $<$ en 2.28 por un \leq . Así, (para el Ejemplo 16.2) tenemos:

$$Pr(1 \leq X \leq 2) = Pr(1 \leq X < 2) = Pr(1 < X \leq 2) = \frac{7}{27}$$

En la figura 2.4 se muestra una gráfica de $p(x)$ para este ejemplo.

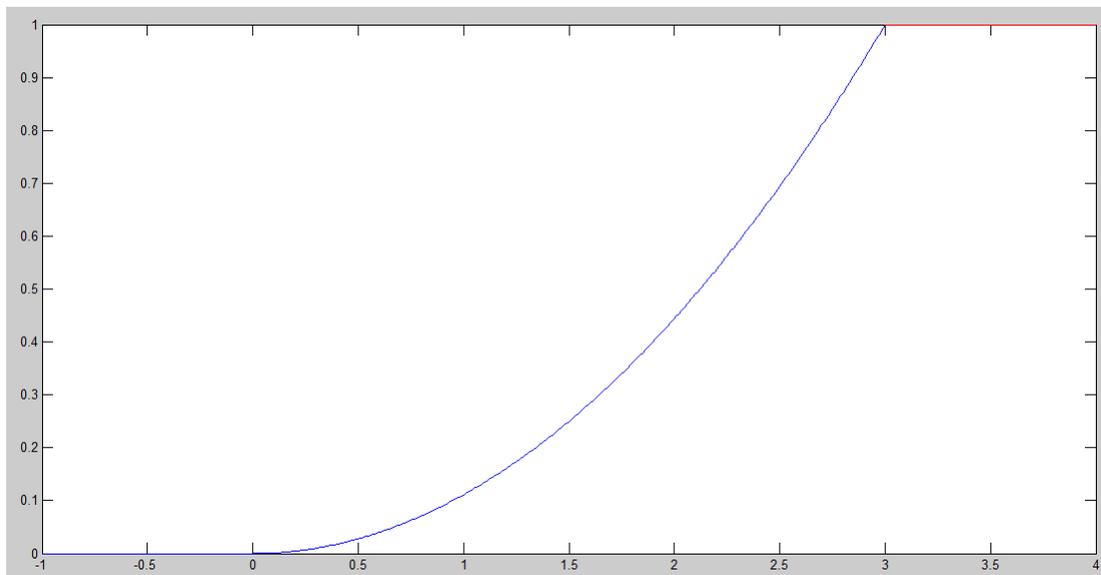


FIGURA 2.4. Función de densidad de probabilidad del ejemplo 16.2

EJEMPLO. 16.3

(a) Encontrar la función de distribución acumulativa, para la variable aleatoria del Ejemplo 16.2 y graficarla.

(b) Usar el resultado de (a) para encontrar $\Pr(1 < X \leq 2)$

Solución para (a): Tenemos de 2.27

$$P(x) = \Pr(X \leq x) = \int_{-\infty}^x p(u) du$$

Si $x < 0$, entonces $P(x) = 0$.

Si $0 \leq x \leq 3$, entonces:

$$P(x) = \int_0^x p(u) du = \frac{x^3}{27}$$

Si $x \geq 3$, tenemos:

$$P(x) = \int_0^3 p(u) du + \int_3^x p(u) du = \int_0^3 \frac{1}{9} u^2 du + \int_3^x 0 du = 1$$

Así la distribución acumulativa requerida es:

$$P(x) = \begin{cases} 0 & x < 0 \\ \frac{x^3}{27} & 0 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

La figura 2.5 muestra la gráfica de $P(x)$.

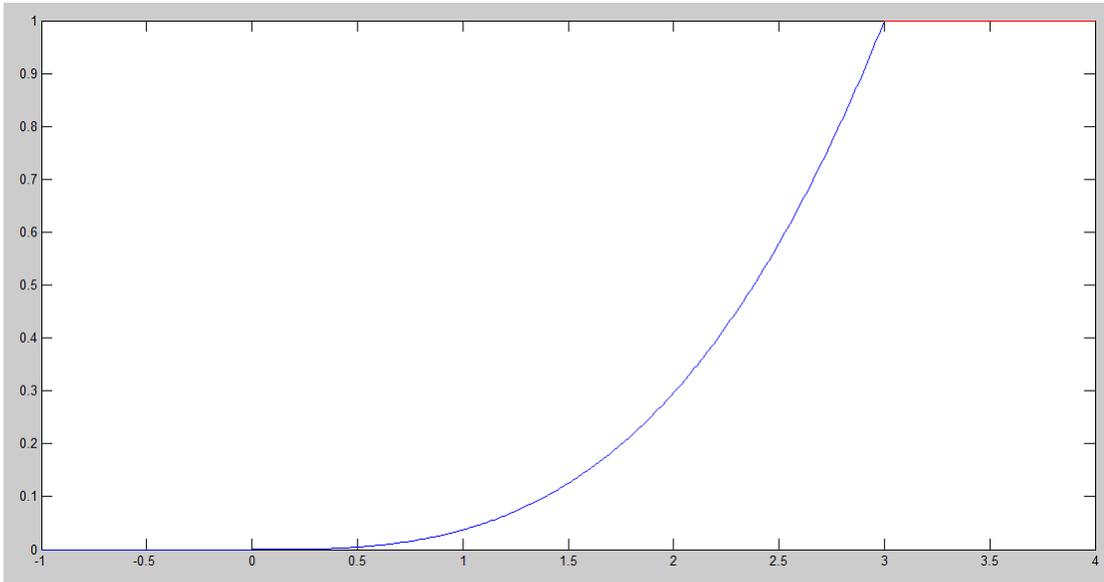


FIGURA 2.5. Función de densidad de probabilidad del ejemplo 16.3

Nótese en la figura 2.5 que $P(x)$ incrementa monótonicamente de 0 a 1 conforme aumenta x . También, se puede observar que a pesar de que $P(x)$ es una función continua a pesar de que $p(x)$ no lo es.

2.17. Obtención de $p(x)$ a partir de $P(x)$

La propiedad básica del teorema fundamental del cálculo

$$\frac{d}{dx} \int_a^x p(u) du = p(x) \quad (2.29)$$

Diferenciando ambos lados de 2.27, es decir,

$$p(x) = \frac{dP(x)}{dx} \quad (2.30)$$

En todos los puntos donde $p(x)$ es continua. Es decir, la derivada de la función de distribución acumulativa es la función de densidad.

2.18. Interpretación gráfica

Si $p(x)$ es la función de densidad de probabilidad para un variable aleatoria X , podemos representar gráficamente $y = p(x)$ como la curva en la Figura 2.6.

Como $p(x) \geq 0$, la curva no puede pasar debajo del eje de las x .

El área entera limitada por la curva y el eje de las x , debe ser igual a 1, debido a la propiedad número dos de las variables aleatorias continuas.

Geoméricamente, la probabilidad de que X esté entre a y b ($Pr(a < X < b)$) está representada por el área sombreada, en la Figura 2.6.

La función de distribución acumulativa $P(x) = Pr(X \leq x)$ es una función monótonicamente creciente, que incrementa de 0 a 1 y se ilustra por la curva de la Figura 2.7.

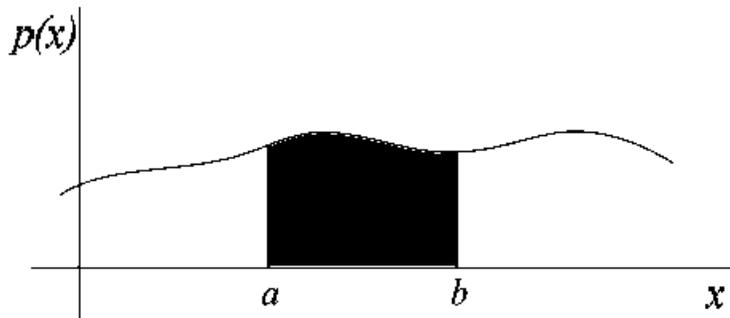


FIGURA 2.6. Se ilustra la función de densidad de probabilidad (pdf)

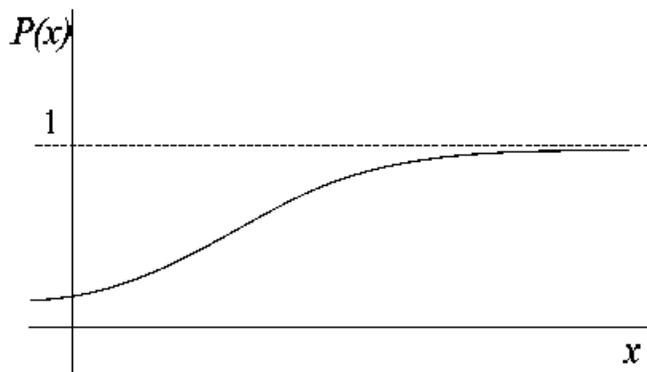


FIGURA 2.7. Se ilustra la función de distribución acumulativa

2.19. Distribuciones de probabilidad conjunta

Los conceptos de función de probabilidad acumulativa (cdf) y la función de densidad de probabilidad (pdf), pueden ser fácilmente generalizadas para dos o mas variables. Consideremos el típico caso de dos variables que sean continuas o discretas ambas. La generalización para tres o mas variables también puede ser hecha.

2.20. Caso de dos variables aleatorias discretas

La función de probabilidad conjunta de dos variables aleatorias discretas X y Y está definida como:

$$p(x, y) = p_{XY}(x, y) = Pr(X = x, Y = y) \quad (2.31)$$

Notamos que en la mayoría de textos técnicos, el subíndice X, Y que indica las variables aleatorias es omitido, como se usó en 2.31

En general $p(x, y)$ es una función de probabilidad si las siguientes propiedades se cumplen:

1. $p(x, y) \geq 0$
2. $\sum_x \sum_y p(x, y) = 1$, donde la suma se hace sobre todos los posibles valores de x , y de todos los valores de y .

Suponiendo que la variable aleatoria X puede tomar cualquiera de los M valores $x_1, x_2, \dots, x_m, \dots, x_M$ y la variable aleatoria Y , puede tomar uno de los N valores $y_1, y_2, \dots, y_n, \dots, y_N$. Así la probabilidad del evento que $X = x_m$ y $Y = y_n$ está dado por:

$$Pr(X = x_m, Y = y_n) = p(x_m, y_n) \geq 0 \quad (2.32)$$

La probabilidad $Pr(X = x_m)$ del evento $X = x_m$ se obtiene por:

$$Pr(X = x_m) = p_X(x_m) = \sum_{n=1}^N p(x_m, y_n) \quad (2.33)$$

Sumando sobre todos los valores de y $\{y_n; n = 1, \dots, N\}$.

Similarmente, la probabilidad $Pr(Y = y_n)$ del evento $Y = y_n$ se obtiene por:

$$Pr(Y = y_n) = p_Y(y_n) = \sum_{m=1}^M p(x_m, y_n) \quad (2.34)$$

Sumando sobre todos los valores de X $\{x_m; m = 1, \dots, M\}$.

En general, las probabilidades $Pr(X = x) = p_X(x)$ y $Pr(Y = y) = p_Y(y)$, se denominan funciones de probabilidad marginal de X y Y . Estas satisfacen las propiedades generales de las funciones de probabilidad:

$$\sum_{n=1}^N p_Y(y_n) = 1 \quad y \quad \sum_{m=1}^M p_X(x_m) = 1 \quad (2.35)$$

Las cuales pueden ser escritas como:

$$\sum_{n=1}^N \sum_{m=1}^M p(x_m, y_n) = 1 \quad (2.36)$$

La probabilidad total siempre es 1.

Función de densidad conjunta (pdf) de dos variables aleatorias continuas X y Y está definida por:

1. $p(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1$

Graficamente, $z=p(x,y)$ representa una superficie, llamada la superficie de probabilidad, como se ilustra en la Figura 2.8. El volumen total limitado por esta superficie y del plano xy es igual a 1, cumpliendo con la propiedad 2 de arriba. La probabilidad de que la variable aleatoria X esté entre a y b, mientras que la variable aleatoria Y esté entre c y d, está dada graficamente por el volumen sombreado de la Figura 2.8 y matemáticamente por

$$Pr(a < X < b, c < Y < d) = \int_{x=a}^b \int_{y=c}^d p(x, y) dx dy \quad (2.37)$$

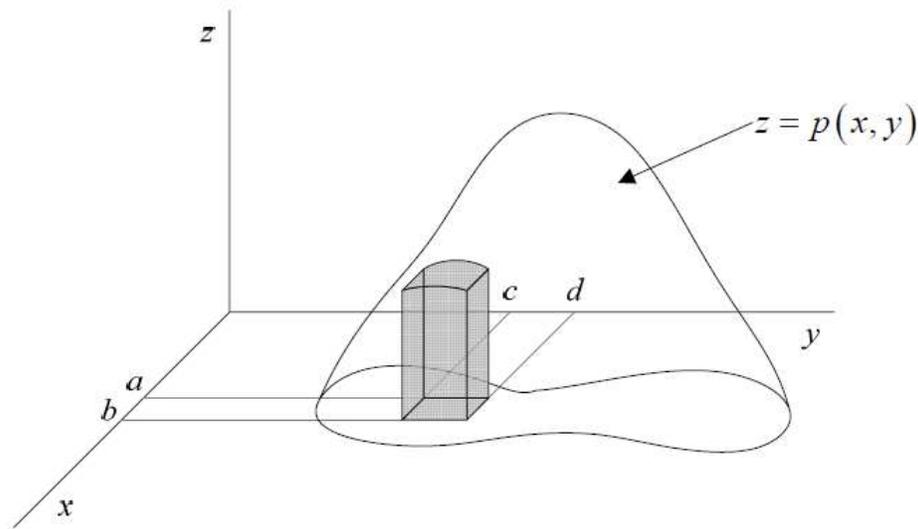


FIGURA 2.8. Se ilustra la superficie de probabilidad y la probabilidad $Pr(a < X < b, c < Y < d)$

En general, si A representa cualquier evento, habrá una región R_A del plano xy que corresponde a él. En dado caso, podemos encontrar la probabilidad de A , efectuando la integración de esta región R_A , así

$$Pr(A) = \int \int_{R_A} p(u, v) du dv \quad (2.38)$$

La función de distribución acumulativa conjunta $P(x, y)$ de X y Y en este caso, está definida por

$$P(x, y) = Pr(X \leq x, Y \leq y) = \int_{u=-\infty}^x \int_{v=-\infty}^y p(u, v) du dv \quad (2.39)$$

La función de densidad de probabilidad conjunta $p(x, y)$ en este caso se obtiene diferenciando la función de distribución acumulativa con respecto a x , y con respecto a y , como sigue

$$p(x, y) = \frac{d^2 P(x, y)}{dx dy} \quad (2.40)$$

Las funciones de distribución acumulativas marginales:

$$P_X(x) = Pr(X \leq x)$$

$$P_Y(y) = Pr(Y \leq y)$$

Y pueden ser obtenidas, integrando la función de densidad conjunta:

$$P_X(x) = Pr(X \leq x) = \int_{u=-\infty}^x \int_{v=-\infty}^{\infty} p(u, v) du dv \quad (2.41)$$

$$P_Y(y) = Pr(Y \leq y) = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^y p(u, v) du dv \quad (2.42)$$

Las funciones de densidad de probabilidad marginales $p_X(x)$ y $p_Y(y)$ son obtenidas integrando:

$$p_X(x) = \int_{v=-\infty}^{\infty} p(x, v) dv \quad y \quad p_Y(y) = \int_{u=-\infty}^{\infty} p(u, y) du \quad (2.43)$$

2.21. Variables aleatorias independientes

Caso discreto. Suponiendo que X y Y son variables aleatorias discretas. Se dice que X y Y son independientes si los eventos $X = x$ y $Y = y$ son eventos

independientes para todo x, y . En este caso

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y) \quad (2.44)$$

Equivalentemente se puede escribir como

$$p(x, y) = p(x)p(y) \quad (2.45)$$

A la inversa, si para todo x , y para todo y , la función de probabilidad conjunta $p(x, y)$ puede ser expresada como un producto de una función que solo depende de x , y una función que solo depende de y (las cuales serían las funciones de probabilidad marginal de X y Y), entonces X y Y son independientes.

Caso continuo. Si X y Y son variables aleatorias continuas, se dice que son independientes si los eventos $X \leq x$ y $Y \leq y$ son eventos independientes para todo x , y para todo y . En ese caso

$$Pr(X \leq x, Y \leq y) = Pr(X \leq x)Pr(Y \leq y) \quad (2.46)$$

Equivalentemente para cdfs

$$P(x, y) = P(x)P(y) \quad (2.47)$$

O equivalentemente para pdfs

$$p(x, y) = p(x)p(y) \quad (2.48)$$

donde $p(x)$ y $p(y)$ son pdfs marginales.

2.22. Distribuciones condicionales

De la sección 7 de este capítulo, sabemos que

$$Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)} \quad (2.49)$$

Si X y Y , son variables aleatorias discretas y tenemos los eventos ($A : X = x$) y ($B : Y = y$), entonces 2.49 se convierte en

$$Pr(Y = y|X = x) = \frac{p(x, y)}{p(x)} = p(B|A) \quad (2.50)$$

donde $p(y|x)$ es llamada función de probabilidad condicional de Y dado X .

Estas ideas, fácilmente, pueden ser extendidas para el caso de las variables continuas aleatorias,

Equivalentemente para la pdf condicional cuando $Y = X$ está dado

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

es decir

$$p(x, y) = p(x|y)p(y) = p(y|x)p(x) \quad (2.51)$$

Esto último resulta ser el tan conocido Teorema de Bayes para la descomposición de una pdf conjunta.

EJEMPLO. 25.1. Usando 2.51 podemos encontrar que la probabilidad de que Y esté entre c y d , dado que $x < X < x + dx$ es

$$Pr(c < Y < d | x < X < x + dx) = \int_c^d p(y|x) dy \quad (2.52)$$

2.23. Esperanza matemática

Un concepto muy importante en probabilidad y estadística es la esperanza matemática, valor esperado, o simplemente esperanza de una variable aleatoria.

■ 1. Caso discreto

Para una variable discreta aleatoria X teniendo los posibles valores $x_1, \dots, x_n, \dots, x_N$, la esperanza de X está definida como:

$$\langle X \rangle = E(X) = x_1 Pr(X = x_1) + \dots + x_N Pr(X = x_N) = \sum_{n=1}^N x_n p(x_n) = \sum xp(x) \quad (2.53)$$

Donde $p(x_n)$ es la función de probabilidad y la suma es tomada de todos los valores de X . En el caso especial de 2.53, donde las probabilidades son iguales, tenemos:

$$\langle X \rangle = E(X) = \frac{x_1 + \dots + x_N}{N} \quad (2.54)$$

la cual es la media aritmética de X .

Si X toma un número infinito de valores, entonces:

$$\langle X \rangle = E(X) = \sum_{n=1}^{\infty} x_n p(x_n) \quad (2.55)$$

donde la serie infinita converge.

■ 2.Caso continuo

Para una variable aleatoria continua X , teniendo la función de densidad $p(x)$, la esperanza de X está definida como:

$$\langle X \rangle = E(X) = \int_{-\infty}^{\infty} xp(x)dx \quad (2.56)$$

donde la integral converge.

La esperanza de X es también llamada la media de X y se denota por m_X o simplemente m cuando se entiende la variable aleatoria en particular.

La media, o esperanza, de X da un sólo valor que actúa como un valor representativo de X , y por ésta razón es llamada medida de tendencia central.

EJEMPLO. 23.1

Supongamos que un juego, es jugado con un solo dado. En éste juego un jugador gana \$20 si el número 2 sale, \$40 si sale el número 4, pierde \$30 si sale el número 6, mientras que el jugador no gana ni pierde si cualquier otra cara del dado sale. Encuentra la suma esperada de dinero que ganaría.

Sea X la variable aleatoria dada por la cantidad de dinero en el lanzamiento. Las cantidades de dinero que ganaría cuando el dado salga 1,2...6 son:

$$x_1 = 0; x_2 = \$20; x_3 = 0; x_4 = \$40; x_5 = 0; x_6 = -\$30$$

Las probabilidades son iguales $p_1 = \dots = p_6 = 1/6$. Por tanto, el valor esperado o la esperanza es:

$$\langle X \rangle = E(X) = (0)\frac{1}{6} + (20)\frac{1}{6} + (0)\frac{1}{6} + (40)\frac{1}{6} + (0)\frac{1}{6} + (-30)\frac{1}{6} = 5$$

Así, el jugador puede esperar ganar \$5.

En un juego justo, por lo tanto, el jugador debe esperar pagar \$5 para jugar el juego.

EJEMPLO. 23.2

La función de densidad de una variable aleatoria continua X está dada por:

$$P(x) = \begin{cases} (1/2)x & 0 < x < 2 \\ 0 & \text{En otro caso} \end{cases}$$

El valor esperado de X es entonces:

$$\langle X \rangle = E(X) = \int_{-\infty}^{\infty} xp(x)dx = \int_0^2 x(\frac{1}{2}x)dx = \int_0^2 \frac{x^2}{2}dx = \frac{x^3}{6} \Big|_0^2 = \frac{8}{6} = \frac{4}{3}$$

2.24. Esperanza de funciones de variables aleatorias

■ 1. Función de una variable aleatoria discreta

Sea X una variable aleatoria discreta con función de probabilidad $p_X(x)$. Entonces

$Y = f(X)$ también es una variable aleatoria discreta, y la función de probabilidad de Y es:

$$p_Y(y) = Pr(Y = y) = \sum_{\{x|f(x)=y\}} Pr(X = x) = \sum_{\{x|f(x)=y\}} p_X(x)$$

Si X toma los valores $x_1, \dots, x_n, \dots, x_N$ y Y toma los valores $y_1, \dots, y_m, \dots, y_M$ ($M \leq N$), entonces:

$$\langle Y \rangle = y_1 p_Y(y_1) + y_2 p_Y(y_2) + \dots + y_M p_Y(y_M) = f(x_1) p_X(x_1) + f(x_2) p_X(x_2) + \dots + f(x_N) p_X(x_N)$$

Por lo tanto

$$\langle f(X) \rangle = \sum_{n=1}^N f(x_n) p_X(x_n) = \sum f(x) p_X(x) \quad (2.57)$$

■ 2. Función de una variable aleatoria continua

De manera similar, si X es una variable aleatoria continua teniendo función de densidad $p_X(x)$, y $Y = f(X)$ entonces la suma 2.57 se transforma en siguiente integral:

$$\langle f(X) \rangle = \int_{-\infty}^{\infty} f(x) p_X(x) dx \quad (2.58)$$

Nótese que 2.57 y 2.58 no involucran, respectivamente, la función de probabilidad y la función de densidad de la variable $Y = f(X)$.

Generalizaciones pueden ser hechas fácilmente para funciones de dos o más variables aleatorias. Por ejemplo si X y Y son dos variables aleatorias continuas que tienen una función de densidad conjunta $p(x, y)$, entonces la esperanza de $f(X, Y)$ está dada por

$$\langle f(X, Y) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) p(x, y) dx dy \quad (2.59)$$

EJEMPLO. 24.1

Considérese X como la variable aleatoria del ejemplo 26.2. Sea $f(X) = 3X^2 - 2X$. Entonces su media será

$$\langle f(X) \rangle = \langle 3X^2 - 2X \rangle = \int_{-\infty}^{\infty} (3x^2 - 2x) p_X(x) dx = \int_0^2 (3x^2 - 2x) \left(\frac{1}{2}x\right) dx = \int_0^2 \frac{3x^3}{2} dx - \int_0^2 x^2 dx = \frac{10}{3}$$

2.25. Teoremas de la esperanza matemática

TEOREMA. 1.

Si c es cualquier constante, entonces

$$\langle cX \rangle = c \langle X \rangle \quad (2.60)$$

TEOREMA. 2.

Si X y Y son cualesquiera variables aleatorias, entonces

$$\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle \quad (2.61)$$

TEOREMA. 3.

Si X y Y son cualesquiera variables aleatorias independientes, entonces

$$\langle XY \rangle = \langle X \rangle \langle Y \rangle \quad (2.62)$$

2.26. La varianza y la desviación estandar

La esperanza de una variable aleatoria X es llamada media y se denota por m . Otra cantidad de gran importancia en probabilidad y estadística es llamada varianza y se define por

$$Var(x) = \sigma_X^2 = \langle (X - m_X)^2 \rangle \quad (2.63)$$

La varianza es un número positivo. La raíz cuadrada positiva de la varianza es llamada la desviación estándar y está dada por

$$\sigma_X = \sqrt{Var(X)} = \sqrt{\langle (X - m)^2 \rangle} \quad (2.64)$$

■ 1. Variable aleatoria discreta

Para una variable aleatoria discreta X teniendo como posibles valores $x_1, \dots, x_n, \dots, x_N$, la varianza de X está dada por

$$\sigma_X^2 = \sum_{n=1}^N (x_n - m)^2 p(x_n) = \sum (x - m)^2 p(x) \quad (2.65)$$

en el caso especial de 2.65 donde las probabilidades son iguales

$$\sigma_X^2 = \left(\frac{1}{N}\right) \sum_{n=1}^N (x_n - m)^2 \quad (2.66)$$

■ 2. Variable aleatoria continua

Para una variable aleatoria continua X teniendo una función de densidad de probabilidad $p(x)$, la varianza de X está dada por

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - m)^2 p(x) dx \quad (2.67)$$

suponiendo que la integral converge.

La varianza (o desviación estándar) es una medida de dispersión, o esparcimiento de valores de la variable aleatoria respecto de la media m . Si los valores tienden a estar concentrados cerca de la media, la varianza es pequeña, mientras que si los valores tienden a estar distribuidos lejos de la media, la varianza es grande. La situación se ilustra gráficamente en la Figura 2.9 para el caso de dos distribuciones continuas teniendo media $m = \mu$.

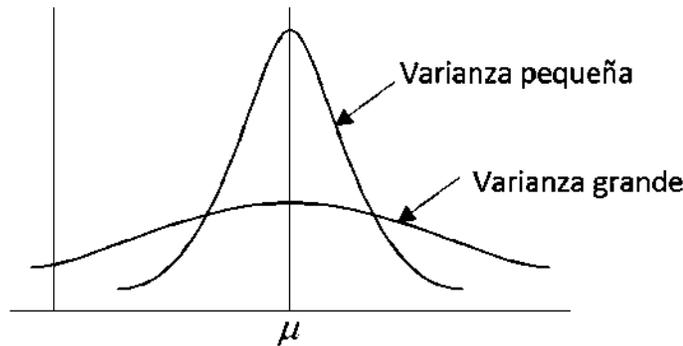


FIGURA 2.9. Ilustrando diferentes varianzas y la misma media

EJEMPLO. 26.1.

Considérese que X es la variable aleatoria del ejemplo 23.2. Su media es $\frac{4}{3}$. Encuentre la varianza.

$$\sigma_X^2 = \langle (X - m)^2 \rangle = \int_{-\infty}^{\infty} (x - \frac{4}{3})^2 p(x) dx = \int_0^2 (x - \frac{4}{3})^2 (\frac{1}{2}x) dx = \frac{2}{9}$$

Nótese que si X tiene ciertas dimensiones o unidades, por ejemplo centímetros (cm), entonces la varianza de X tiene unidades cm^2 mientras que la desviación estándar tiene las mismas unidades que X . Es por ésta razón que la desviación estándar es muy utilizada.

2.27. Teoremas de la varianza

TEOREMA. 1.

Si m es el valor medio de X , entonces

$$\sigma_X^2 = \langle (X - m)^2 \rangle = \langle X^2 \rangle - m^2 = \langle X^2 \rangle - \langle X \rangle^2 \quad (2.68)$$

TEOREMA. 2.

Si c es cualquier constante, entonces

$$\text{Var}(cX) = c^2 \text{Var}(X) \quad (2.69)$$

TEOREMA. 3.

La cantidad $\langle (X - a)^2 \rangle$ es mínima cuando $a = m = \langle X \rangle$

TEOREMA. 7.

Si X y Y son cualesquiera variables aleatorias independientes, entonces:

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \quad (2.70)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) \quad (2.71)$$

2.28. Variables aleatorias estandarizadas

Sea X una variable aleatoria con media m y una desviación estándar σ . Entonces podemos definir una variable estandarizada asociada a X , dada por

$$\hat{X} = \frac{X - m}{\sigma} \quad (2.72)$$

Una propiedad importante de \hat{X} es que tiene una media cero y una varianza 1, por ello recibe el nombre de estandarizada.

$$\langle \hat{X} \rangle = 0, \text{Var}(\hat{X}) = 1 \quad (2.73)$$

A los valores de una variable estandarizada se les conoce también como valores estándar, y X se dice que se expresa en unidades estándar. (Por ejemplo, σ se toma como las unidades en que se mide la diferencia $X - m$)

2.29. Momentos

El n -ésimo momento de una variable aleatoria X respecto de la media m , también llamado el n -ésimo momento central, está definido como:

$$m_n = \langle (X - m)^n \rangle \quad (2.74)$$

donde $n = 0, 1, 2, \dots$

Así, por ejemplo

$$m_0 = 1, m_1 = 0, m_2 = \sigma^2 = \text{Var}(X) \quad (2.75)$$

Asumiendo convergencia absoluta, el n -ésimo momento central se calcula como sigue:

$$m_n = \sum (x - m)^n p(x), \quad (\text{Variable discreta}) \quad (2.76)$$

$$m_n = \int_{-\infty}^{\infty} (x - m)^n p(x) dx, \quad (\text{Variable continua}) \quad (2.77)$$

El n -ésimo momento de una variable aleatoria X respecto del origen, también llamado el n -ésimo momento crudo, está definido como

$$m'_n = \langle X^n \rangle \quad (2.78)$$

La relación entre los momentos central y crudo está dada por

$$m_n = m'_n - \binom{n}{1} m'_{n-1} + \dots + (-1)^j \binom{n}{j} m'_{n-j} m^j + \dots + (-1)^n m'_0 m^n \quad (2.79)$$

Como caso especial tenemos, usando $m'_1 = m$ y $m'_0 = 1$,

$$m_2 = m'_2 - m^2; m_3 = m'_3 - 3m'_2 m + 2m^2 \quad (2.80)$$

2.30. Función generadora de momentos

La función generadora de momentos de X está definida por

$$M_X(t) = \langle e^{tX} \rangle \quad (2.81)$$

Así, asumiendo convergencia,

$$M_X(t) = \sum e^{tX} p(x), \quad (\text{Variable discreta}) \quad (2.82)$$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tX} p(x) dx, \quad (\text{Variable continua}) \quad (2.83)$$

Se puede mostrar que la expansión de serie de Taylor para la función generadora de momentos es la siguiente:

$$M_X(t) = 1 + mt + m_2' \frac{t^2}{2!} + \dots + m_n' \frac{t^n}{n!} + \dots \quad (2.84)$$

Como los coeficientes en ésta expansión nos permiten encontrar los momentos, la razón de que se llame función generadora de momentos es clara, por ejemplo:

$$m_n' = \frac{d^n}{dt^n} M_X(t) \Big|_{t=0} \quad (2.85)$$

Es decir, m_n' es la n - ésima derivada de $M_X(t)$ evaluada en $t = 0$.

2.31. Media, varianza y covarianza para distribuciones conjuntas

Los resultados obtenidos arriba pueden ser extendidos para dos o mas variables. Por ejemplo, si X y Y son variables aleatorias teniendo funciones de densidad conjunta $p_{XY}(x, y) = p(x, y)$, las medias de X y Y son:

$$m_X = \langle X \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x, y) dx dy; \quad m_Y = \langle Y \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yp(x, y) dx dy \quad (2.86)$$

y las varianzas son:

$$\sigma_X^2 = \langle (X - m_X)^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_X)^2 p(x, y) dx dy \quad (2.87)$$

$$\sigma_Y^2 = \langle (Y - m_Y)^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - m_Y)^2 p(x, y) dx dy \quad (2.88)$$

Otra cantidad que surge en el caso de dos variables X y Y es la covarianza, definida por:

$$\sigma_{XY} = Cov(X, Y) = \langle (X - m_X)(Y - m_Y) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)(y - m_y) p(x, y) dx dy \quad (2.89)$$

Se hace lo mismo para el caso de dos variables aleatorias discretas.

2.32. Teoremas de la covarianza

TEOREMA. 1.

$$\sigma_{XY} = Cov(X, Y) = \langle (X - m_X)(Y - m_Y) \rangle = \langle XY \rangle - m_X m_Y \quad (2.90)$$

TEOREMA. 2.

Si X y Y son variables aleatorias independientes, entonces:

$$\sigma_{XY} = Cov(X, Y) = 0 \quad (2.91)$$

TEOREMA. 3.

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y) \quad (2.92)$$

TEOREMA. 4.

$$|\sigma_{XY}| \leq \sigma_X \sigma_Y \quad (2.93)$$

2.33. Coeficiente de correlación

Si X y Y son variables aleatorias independientes, entonces $\sigma_{XY} = Cov(X, Y) = 0$. Por otro lado, si X y Y son completamente dependientes, por ejemplo $X = Y$, entonces $Cov(X, Y) = \sigma_{XY} = \sigma_X \sigma_Y$. Ésto nos lleva a medir la dependencia de las variables X y Y , dado por un coeficiente ρ que se define como sigue

$$\rho = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}; \quad -1 \leq \rho \leq 1 \quad (2.94)$$

Llamamos a ρ el coeficiente de correlación. En el caso donde $\rho = 0$ (Por ejemplo, que la covarianza sea cero) llamamos a las variables X y Y no correlacionadas.

2.34. Otras medidas de tendencia central

Mediana. La mediana es el valor para el cual $Pr(X < x) = 1/2 = Pr(X > x)$.

Percentiles. A veces es conveniente subdividir el área bajo la curva de una función de densidad usando alguna abscisa, para que el área que quede a la izquierda de la abscisa sea algún porcentaje del total del área unitaria. Los valores correspondientes a tales áreas se llaman percentiles. La mediana sería el 50^{avo} percentil.

2.35. Desviación media

La desviación media (D.M) de una variable aleatoria X está definida como la esperanza de $|X - m|$, así, asumiendo convergencia:

$$D.M.(X) = \langle |X - m| \rangle = \sum |x - m| p(x), \quad (\text{Variable discreta}) \quad (2.95)$$

$$D.M.(X) = \langle |X - m| \rangle = \int_{-\infty}^{\infty} |x - m| p(x) dx, \quad (\text{Variable continua}) \quad (2.96)$$

2.36. Sesgo

Algunas veces una distribución no es simétrica cerca de ningún valor, pero en su lugar tiene un valor donde una de sus “colas” es mas larga que la otra, como se ilustra en la figura 2.10 y figura 2.11. Las medidas que describen ésta asimetría son llamadas coeficientes de sesgado, o sólo sesgado. Ésta medida está dada por

$$\alpha_3 = \frac{m_3}{\sigma^3} \quad (2.97)$$

La medida α_3 será positiva o negativa de acuerdo a si la distribución está sesgada a la izquierda o a la derecha, respectivamente. Para una distribución simétrica, $\alpha_3 = 0$

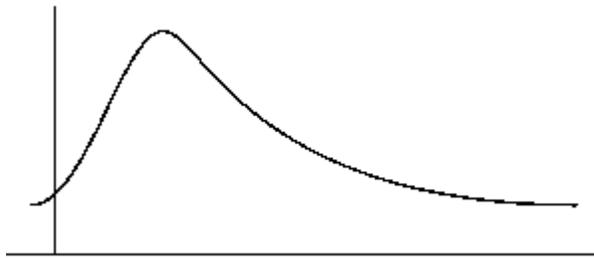


FIGURA 2.10. Sesgado a la derecha

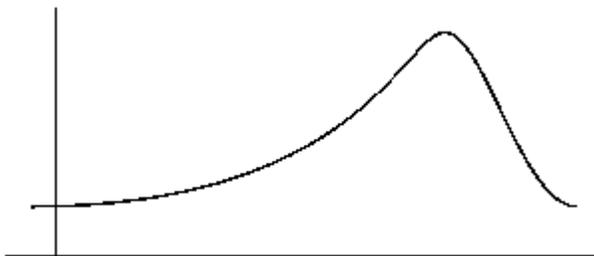


FIGURA 2.11. Sesgado a la izquierda

2.37. Kurtosis

En algunos casos una distribución puede tener sus valores concentrados cerca de la media, así las distribuciones tienen un “pico” largo como el indicado en la curva de la figura 2.12. En algunos otros casos, la distribución puede ser relativamente plana como la curva punteada en la figura 2.12. Las medidas del grado de “picudez” de una distribución son llamados coeficientes de kurtosis, o sólo kurtosis. Una medida muy usada, la cual está dada por

$$\alpha_4 = \frac{m_4}{\sigma^4} \quad (2.98)$$

Éste coeficiente es usualmente comparado con la curva de función de densidad normal estándar, que tiene una kurtosis igual a 3.

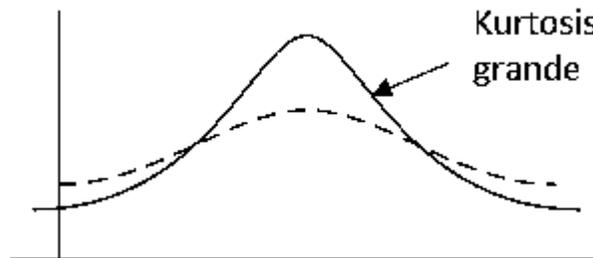


FIGURA 2.12. Kurtosis

2.38. Distribución binomial

En muchos experimentos (como lanzar un dado o una moneda) la probabilidad de hacer muchos intentos, no cambia de un intento a otro. Éstos intentos se dice que son independientes y son llamados ensayos de Bernoulli.

Sea p la probabilidad de éxito, o sea, la probabilidad de que un evento suceda en un sólo intento. Entonces $q = 1 - p$ es la probabilidad de fallar, es decir, la probabilidad de que el evento fallará en un sólo intento. Entonces la probabilidad de que el evento ocurra exactamente x veces en N intentos, está dada por la función de probabilidad siguiente (distribución binomial)

$$p(x) = Pr(X = x) = \binom{N}{x} p^x q^{N-x} = \frac{N!}{x!(N-x)!} p^x q^{N-x} \quad (2.99)$$

donde la variable aleatoria X denota el número de éxitos en N intentos y $x = 0, 1, \dots, N$.

EJEMPLO. 38.1

La probabilidad de obtener exactamente 2 caras en 6 lanzamientos de una moneda es

$$Pr(X = 2) = \binom{6}{2} p^2 q^{6-2} = \frac{6!}{2!(6-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = \frac{15}{64}$$

El nombre “distribución binomial” es usado debido a que la expresión 2.99, corresponde a los términos sucesivos de una expansión binomial.

$$(p + q)^N = \sum_{x=0}^N \binom{N}{x} p^x q^{N-x} \quad (2.100)$$

Algunas de las propiedades importantes de la distribución binomial están listadas en la siguiente tabla:

Media	$m = Np$
Varianza	$\sigma^2 = Npq$
Desviación estandar	$\sigma = \sqrt{Npq}$
Sesgado	$\alpha_3 = \frac{m_3}{\sigma^3} = \frac{q-p}{\sqrt{Npq}}$
Kurtosis	$\alpha_4 = \frac{m_4}{\sigma^4} = 3 + \frac{1-6pqm_4}{Npq}$
Función generadora de momentos	$M(t) = (q + pe^t)^N$

2.39. Distribución normal

Uno de los más importantes ejemplos de una distribución de probabilidad continua es la distribución normal (Gaussiana), con la función de densidad dada por

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \quad (2.101)$$

donde m y σ son la media y la desviación estándar, respectivamente. Si X es una variable aleatoria que tiene distribución normal dada por 2.101 podemos denotarla como $X \sim N(m, \sigma^2)$.

La función de distribución acumulativa está dada por

$$P(x) = Pr(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(v-m)^2}{2\sigma^2}\right) dv \quad (2.102)$$

Introduciremos una variable estandarizada Z correspondiente a X , es decir

$$Z = \frac{X - m}{\sigma} \quad (2.103)$$

La función de densidad normal estándar para Z , puede ser fácilmente obtenida de 2.101 haciendo $m = 0$ y $\sigma = 1$, así

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \quad (2.104)$$

La función de distribución acumulativa correspondiente está dada por:

$$P(z) = Pr(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{u^2}{2}\right) du = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^z \exp\left(-\frac{u^2}{2}\right) du = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right) \quad (2.105)$$

donde

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-u^2) du \quad (2.106)$$

también llamada función del error, y escribimos $Z \sim N(0, 1)$.

Una gráfica de la función de densidad 2.104 (llamada curva normal estándar), es mostrada en la Figura 2.13. En ésta gráfica indicamos el área entre 1, 2 y 3 desviaciones estándar respecto a la media (por ejemplo, entre $z = -1$ y $+1$, $z = -2$ y $+2$, $z = -3$ y $+3$) respectivamente, para el 68.27%, 95.45% y 99.73 del área total, que es uno. Ésto significa:

$$Pr(-1 \leq Z \leq 1) = 0.6827,$$

$$Pr(-2 \leq Z \leq 2) = 0.9545,$$

$$Pr(-3 \leq Z \leq 3) = 0.9973.$$

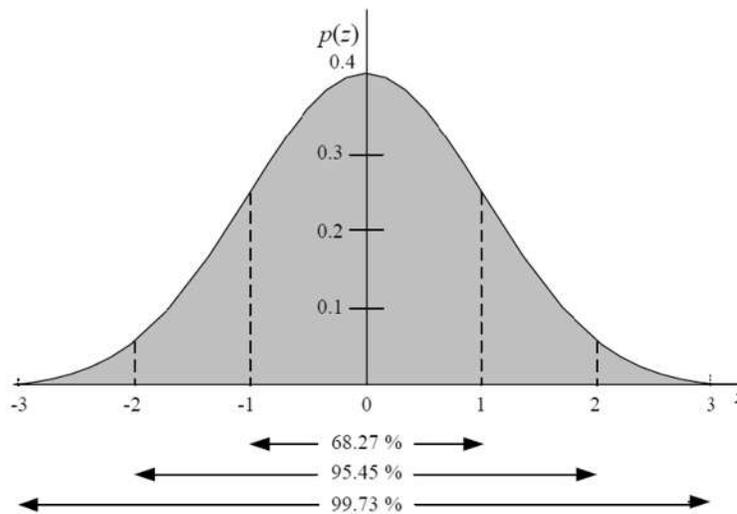


FIGURA 2.13. Curva normal estándar

Algunas de las propiedades importantes de la función de distribución se enuncian en la siguiente tabla:

Media	m
Varianza	σ^2
Desviación estandar	σ
Sesgado	$\alpha_3 = 0$
Kurtosis	$\alpha_4 = 3$
Función generadora de momentos	$M(t) = \exp(mt + \frac{\sigma^2 t^2}{2})$

2.40. Procesos estocásticos

Un proceso estocástico es un concepto matemático que sirve para caracterizar una sucesión de variables aleatorias (estocásticas) que evolucionan en función de otra variable, generalmente el tiempo y se denota por $X(t)$. Entonces, se puede ver un proceso estocástico como un conjunto de funciones en el dominio del tiempo que son los resultados de cierto experimento. Como la cdf y la pdf de un proceso estocástico son dependientes del tiempo, se denotarán como $P_X(x; t)$ y $p_X(x; t)$. El n -ésimo momento de un proceso estocástico $X(t)$ es:

$$m_X^n(t) = \langle X^n(t) \rangle = \int_{-\infty}^{\infty} x^n p_X(x, t) dx \quad (2.107)$$

En los procesos estocásticos existe la posibilidad de que sean continuos, discretos o continuo y discreto, en el cuadro 2.2 se observa esto:

	t Discreto	t Continuo
X Discreta	Proceso de estado discreto, tiempo discreto	Proceso de estado discreto, tiempo continuo
X Continua	Proceso de estado continuo, tiempo discreto	Proceso de estado continuo, tiempo continuo

CUADRO 2.2. Tipos de procesos estocásticos (Continuos y discretos)

2.41. Ruido blanco gaussiano

El ruido blanco gaussiano es un proceso estocástico que tiene de una función de distribución normal, con una media cero $m = 0$ y una varianza σ^2 .

Su función de correlación de la forma δ (Impulso unitario) está dada por

$$R_n(\tau) = N_0\delta(\tau) \quad (2.108)$$

Donde N_0 es la densidad espectral del ruido.

Debido a la forma δ de la función de correlación, la función de densidad espectral $P_n(f)$ del ruido blanco, es una constante. Entonces se tiene

$$P_n(f) = \mathcal{F}\{R_n(\tau)\} = \int_{-\infty}^{\infty} R_n(\tau)\exp(-i2\pi f\tau)d\tau = N_0 \quad (2.109)$$

para todas las frecuencias de $-\infty < f < \infty$, entonces la energía total del ruido blanco es infinita. Este proceso es un idealización de un proceso de ruido real que ocurre en mediciones eléctricas, electrónicas o comunicaciones y algunos otros sistemas de procesamiento de señales.

Capítulo 3

Filtro de Kalman

3.1. Filtro de kalman discreto

En este capítulo se presenta el algoritmo del filtro de Kalman de acuerdo a su versión original, Kalman (1960) donde las mediciones ocurren y el estado es estimado en puntos discretos en el tiempo. La mayor parte del material presentado en este capítulo ha sido tomado de Bishop and Welch (2001)

3.2. El proceso de la estimación

El filtro de Kalman está dirigido al problema general de estimar el estado $x_k \in \mathfrak{R}^n$ de un proceso controlado en tiempo discreto que es gobernado por una ecuación de diferencias estocástica lineal de la forma siguiente, llamada ecuación de estado

$$x_k = Ax_{k-1} + Bu_k + w_{k-1} \quad (3.1)$$

con una medición $z \in \mathfrak{R}^m$ dada por la ecuación de salida presente, donde x_k es el vector de estado en el instante presente k y x_{k-1} es el estado anterior.

$$z_k = Hx_k + v_k \quad (3.2)$$

Las variables aleatorias w_k y v_k representan el ruido del proceso y el ruido de la medición respectivamente. Se asumen independientes (el uno del otro), ruido blanco y con distribuciones de probabilidad normales, es decir,

$$p(w) \sim N(0, Q) \quad (3.3)$$

$$p(v) \sim N(0, R) \quad (3.4)$$

En la práctica, las matrices de la covarianza del ruido del proceso Q y la covarianza de ruido de la medición R , pueden cambiar en cada instante de tiempo o de medición, sin embargo aquí supondremos que son constantes.

La matriz A de dimensión $n \times n$, en la ecuación de diferencias 3.1 relaciona el estado en el instante anterior de tiempo $k - 1$ con el estado del instante de tiempo actual k . Nótese que en la práctica la matriz A puede cambiar de un instante de tiempo a otro, pero supondremos que es constante. La matriz B de dimensión $n \times l$, relaciona la entrada de control $u \in \mathfrak{R}^l$ al estado x . La matriz H de dimensión $m \times n$ en la ecuación de medición 3.2 relaciona el estado con la medición z_k . En la práctica H puede cambiar con cada instante de tiempo o medición, pero supondremos que es constante.

3.3. La estrategia computacional del filtro de Kalman

Definimos $\hat{x}_k^- \in \mathfrak{R}^n$ (Nótese el super índice negativo) como nuestro estado estimado *a priori* en el instante k , dándonos información anterior del proceso al instante k , y $\hat{x}_k \in \mathfrak{R}^n$ será nuestra estimación del estado *a posteriori* (o corregido) al instante k dada la medición z_k . Podemos entonces definir los errores de las estimaciones a priori y a posteriori como sigue

$$e_k^- = x_k - \hat{x}_k^- \text{ y } e_k = x_k - \hat{x}_k \quad (3.5)$$

La covarianza a priori de la estimación del error es entonces

$$P_k^- = E[e_k^- e_k^{-T}] \quad (3.6)$$

y la covarianza a posteriori de la estimación del error es

$$P_k = E[e_k e_k^T] \quad (3.7)$$

En las ecuaciones 3.6 y 3.7 y de aquí en adelante se usará la notación $E[\]$ para representar el operador esperanza matemática, denotado por $\langle \ \rangle$ en el capítulo anterior.

En lo que corresponde a las ecuaciones para el filtro de Kalman, comenzamos con la meta de encontrar la ecuación que calcule un estado a posteriori estimado \hat{x}_k como una combinación lineal de la estimación del estado a priori \hat{x}_k^- y una ponderación de la diferencia entre la medición actual z_k y la predicción de la medición basada en la estimación a priori de estado $H\hat{x}_k^-$ como se muestra en la ecuación 3.8. Las justificaciones de la ecuación 3.8 están dadas en Brown et al. (1992).

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \quad (3.8)$$

La diferencia $(z_k - H\hat{x}_k^-)$ en la ecuación 3.8 es llamada la “inovación de la medición”, o “residuo”. El residuo refleja la discrepancia entre la medición predicha $H\hat{x}_k^-$ y la medición actual z_k . Un residuo igual a cero, significa que las dos son completamente iguales.

La matriz K de dimensión $n \times m$ 3.8 es escogida para ser la ganancia que minimice la covarianza a posteriori del error dado por 3.7. Esta minimización puede ser llevada a cabo de la siguiente forma, primero sustituyendo 3.8 en la definición anterior de e_k , sustituyendolo en 3.7, calculando las esperanzas indicadas, sacando la derivada la ecuación que resulta con respecto a K , igualando el resultado a cero y resolviendo para K . Para mas detalles ver Ham and Brown (1983); Jacobs and Jacobs (1993); Peter (1979). Una forma de escribir K (Hay muchas formas pero ésta es una de las mas populares) que minimiza la ecuación 3.7 está dada por

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \quad (3.9)$$

Para el caso donde $(H P_k^- H^T + R)^{-1}$ es una constante, la ecuación de la ganancia de Kalman se puede reescribir como:

$$K_k = \frac{P_k^- H^T}{(H P_k^- H^T + R)^{-1}} \quad (3.10)$$

De la ecuación 3.9 podemos ver que mientras la covarianza del ruido de medición R se aproxima a cero, la ganancia K hace que el residuo tenga más peso y la corrección de la estimación a priori sea mayor, específicamente:

$$\lim_{R_k \rightarrow 0} K_k = H^{-1} \quad (3.11)$$

Por otro lado, mientras que la covarianza a priori de la estimación del error P_k^- se aproxima a cero, la ganancia K hace que el residuo sea menos pesado y la corrección de la estimación a priori sea menor. específicamente:

$$\lim_{P_k^- \rightarrow 0} K_k = 0 \quad (3.12)$$

Otra forma de pensar acerca del peso que se le da al residuo por medio de la ganancia K , es que mientras que la covarianza del ruido de medición R se aproxima a cero, la medición actual z_k es mas confiable, mientras que la predicción de la medición $H\hat{x}_k^-$ es menos confiable. Por otro lado, mientras que la covarianza a priori de la estimación del error P_k^- se aproxima a cero, la medición actual z_k es menos confiable y la medición predicha $H\hat{x}_k^-$ es más confiable.

3.4. Las propiedades probabilísticas del filtro de Kalman

El filtro de Kalman tiene la propiedad de conservar los primeros dos momentos de la distribución del estado, es decir,

$$E[x] = \hat{x}_k \quad (3.13)$$

$$E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] = P_k \quad (3.14)$$

La estimación a posteriori del estado dada por la ecuación 3.8 refleja la media (el primer momento) de la función de distribución. Está distribuido de forma normal si las condiciones 3.3 y 3.4 se satisfacen. La covarianza estimada del error a posteriori 3.7 refleja la covarianza de la distribución del estado (el segundo momento central). En otras palabras:

$$p(x_k | z_k) \sim N(E[x_k], E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]) \quad (3.15)$$

$$p(x_k | z_k) \sim N(\hat{x}_k, P_k) \quad (3.16)$$

Para más detalles sobre estas propiedades probabilísticas del filtro de Kalman, ver Ham and Brown (1983)

3.5. El algoritmo discreto del filtro de Kalman

El filtro de Kalman estima un proceso usando una forma de retroalimentación, el filtro estima el estado del proceso en algún instante de tiempo y entonces obtiene una retroalimentación a partir de las mediciones (ruidosas) de las salidas del proceso. Así, las ecuaciones para el filtro de Kalman caen en dos grupos: ecuaciones de actualización de tiempo y ecuaciones de actualización basadas en la medición. Las ecuaciones de actualización de tiempo son las responsables de proyectar hacia delante (en el tiempo) una estimación el estado actual y de la covarianza del error, para obtener una estimación a priori para el siguiente instante de tiempo. Las ecuaciones de actualización de la medición son responsables de la retroalimentación, que permite incorporar una nueva medición en la estimación a priori para obtener un estimación a posteriori mejorada.

Las ecuaciones de actualización en el tiempo, se pueden pensar como las ecuaciones de predicción, mientras que las ecuaciones de la actualización de la medición pueden ser pensadas como ecuaciones de corrección. De hecho, el algoritmo final de

estimación se parece a un algoritmo predictor-corrector para resolver problemas numéricos como se muestra en la Figura 3.1

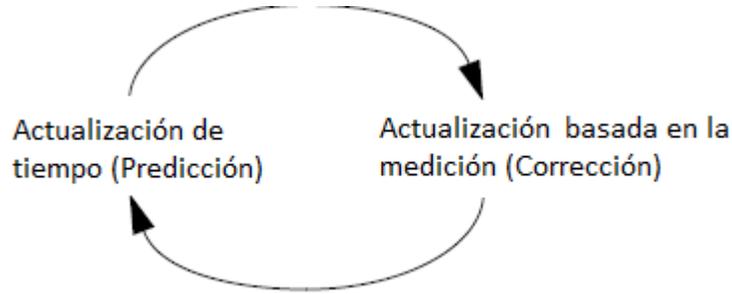


FIGURA 3.1. Etapas de predicción y corrección del filtro de Kalman

Las ecuaciones para la actualización de tiempo son las siguientes:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k \quad (3.17)$$

$$P_k^- = AP_{k-1}A^T + Q \quad (3.18)$$

De nuevo notamos que las ecuaciones de actualización de arriba proyectan la estimación del estado y de la covarianza hacia delante de un instante de tiempo $k - 1$ hacia el instante k . A y B son las matrices de la ecuación de estado de 3.1, mientras que Q es la covarianza del proceso de 3.3.

Las ecuaciones para la actualización basada en la medición son las siguientes:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad (3.19)$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \quad (3.20)$$

$$P_k = (I - K_k H)P_k^- \quad (3.21)$$

La primera tarea durante la actualización basada en la medición es calcular la ganancia de Kalman, K_k . Nótese que la ecuación dada en 3.19 es la misma que 3.9. El siguiente paso es medir el proceso para obtener z_k , y entonces generar la estimación del estado a posteriori incorporando la medición en 3.20. De nuevo 3.20 es la misma que 3.8 (se repite para que las ecuaciones de actualización del tiempo queden completas). El paso final es obtener una covarianza a posteriori estimada del error mediante 3.21.

Después de cada par de actualizaciones en el tiempo y en la medición, el proceso se repite con la estimación a posteriori previa y es usada para proyectar o predecir una nueva estimación a priori. La naturaleza recursiva es uno de los atractivos del filtro de Kalman. La figura 3.2 ofrece una imagen completa de la operación del filtro, combinando el diagrama de alto nivel de la figura 3.1 con las ecuaciones de actualización de tiempo y de medición. Bishop and Welch (2001)

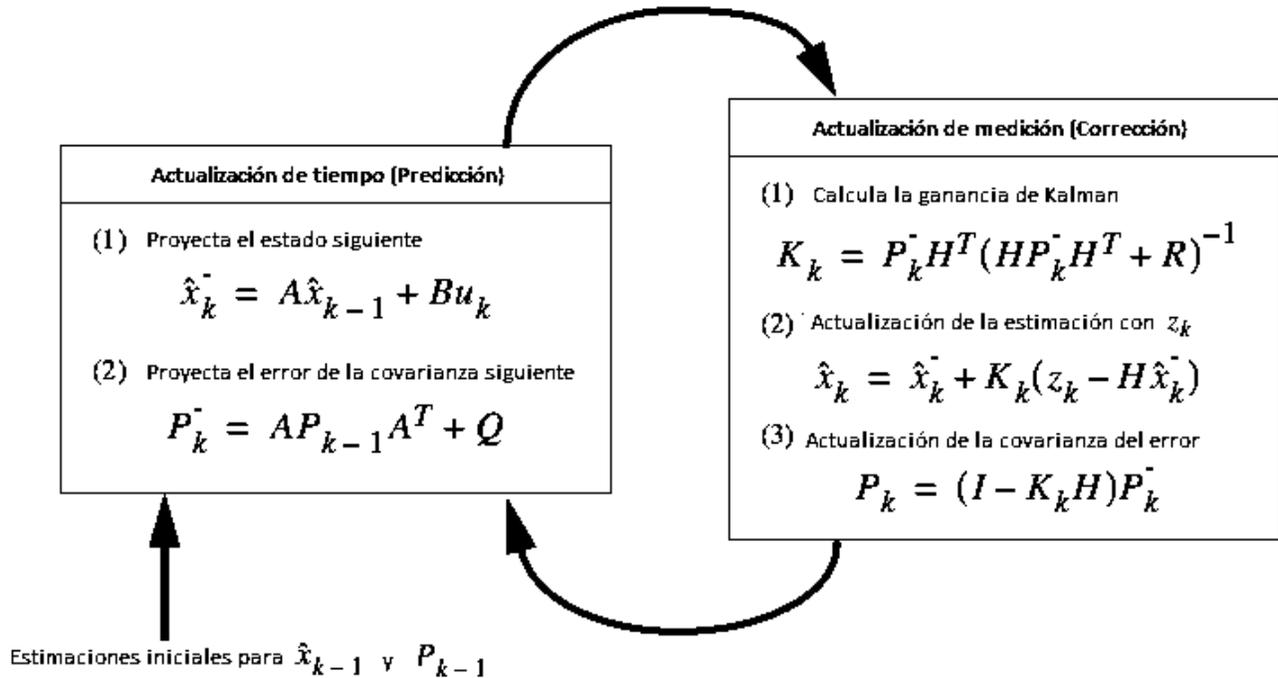


FIGURA 3.2. Imagen completa de la operación del filtro de Kalman

3.6. El filtro de Kalman como observador de estados

El filtro de Kalman se puede ver también como un observador de estados óptimo en el sentido de que minimiza la covarianza del error de estimación. El diagrama de bloques del observador de estados tipo Luenberger es muestra en la figura 3.3

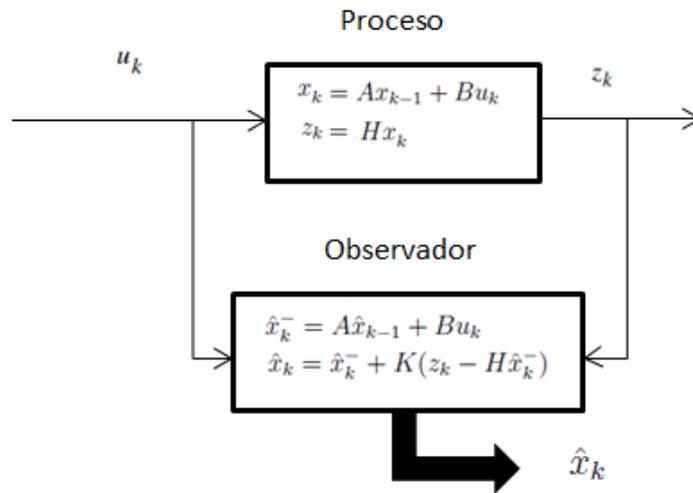


FIGURA 3.3. Observador de estados tipo Luenberger

El algoritmo de estimación que se ejecuta en un observador de Luenberger se puede escribir como sigue:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k \quad (3.22)$$

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \quad (3.23)$$

Donde K es una matriz constante que se diseña para conseguir la convergencia en lazo cerrado del observador.

Como puede observarse, las ecuaciones 3.22, 3.23 son idénticas a las usadas en el filtro de Kalman, excepto que en el observador de Luenberger la matriz K es constante, mientras que en el filtro de Kalman la correspondiente matriz K_k varía en cada instante de muestreo para garantizar la minimización de la covarianza del error de estimación.

Es decir, el filtro de Kalman no es más que un observador tipo Luenberger, al cual se le ha agregado un algoritmo de optimización de la matriz de covarianza P_k , esto se puede ilustrar si representamos el filtro de Kalman por su diagrama de bloques como en la figura 3.4

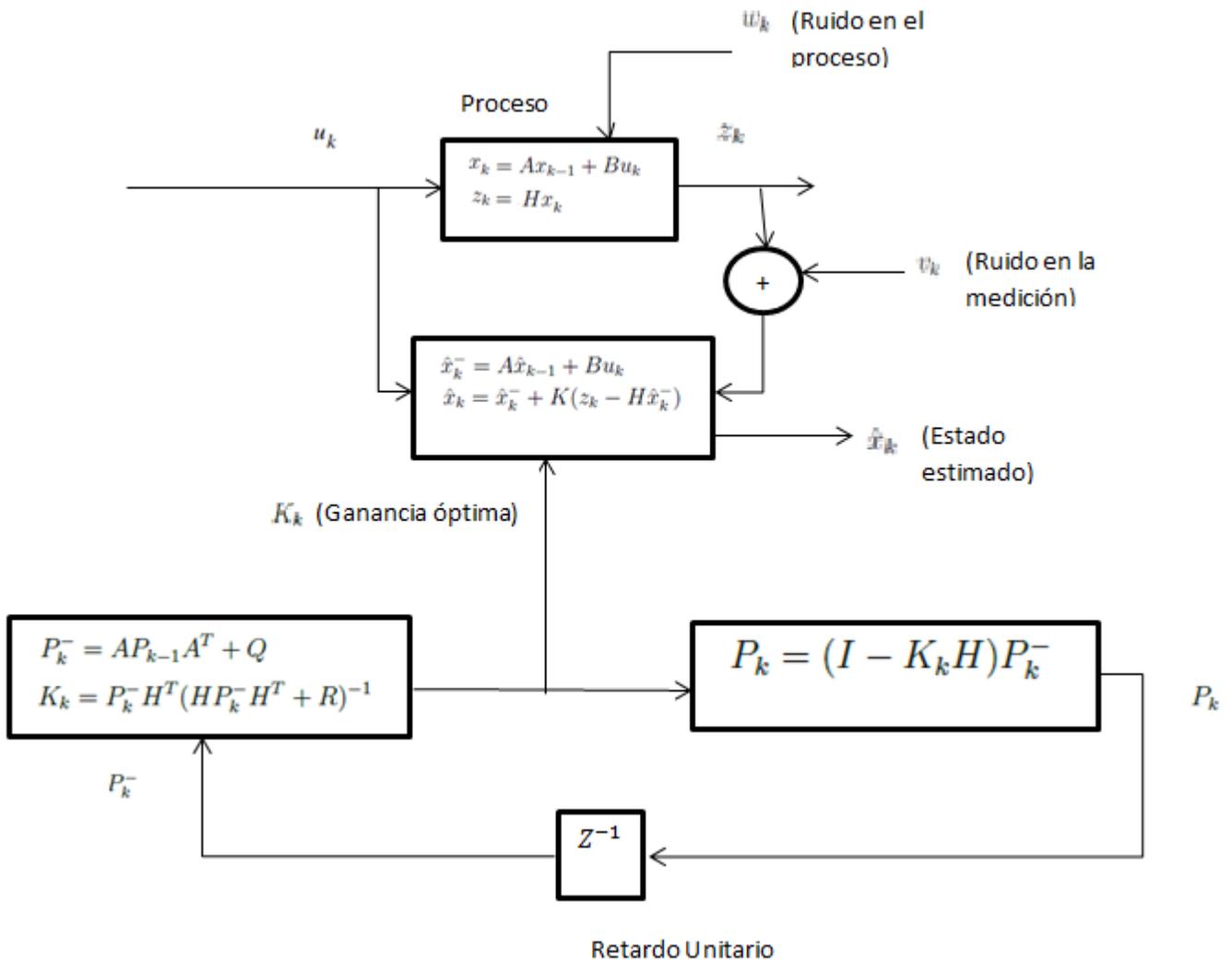


FIGURA 3.4. El filtro de Kalman como un observador de Luenberger más optimizador de la ganancia K_k

3.7. Ejemplo Sistema de nivel del líquido

Para ejemplificar el filtro de Kalman, se tomará en cuenta un sistema de nivel de líquido, que está construido como se muestra en la figura 3.5.

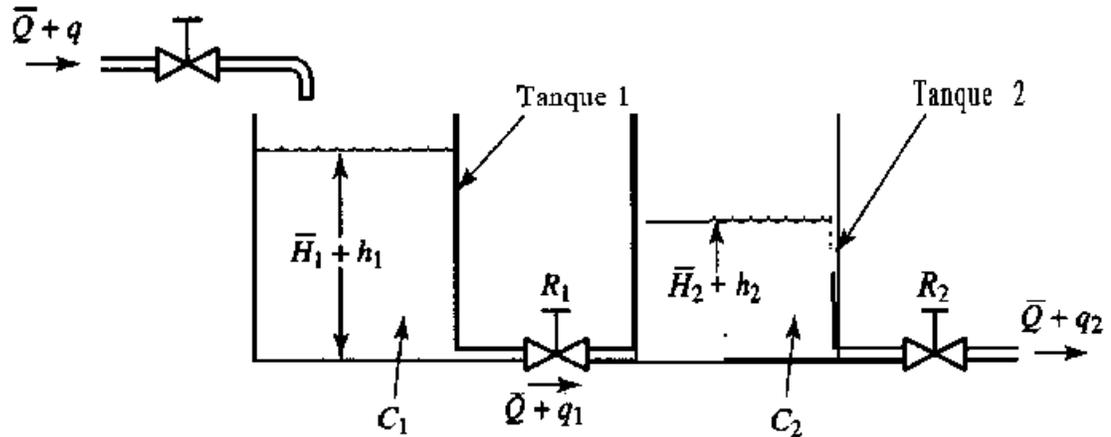


FIGURA 3.5. Sistema de nivel de líquido

Donde:

\bar{Q} = Velocidad de flujo en estado estable

\bar{H}_1 = Nivel del líquido en estado estable del tanque 1

\bar{H}_2 = Nivel del líquido en estado estable del tanque 2

El modelo del sistema de nivel de líquido está dada por. (ver Ogata (2003)) las ecuaciones siguientes:

$$\frac{h_1 - h_2}{R_1} = q_1 \quad (3.24)$$

$$C_1 \frac{dh_1}{dt} = q - q_1 \quad (3.25)$$

$$\frac{h_2}{R_2} = q_2 \quad (3.26)$$

$$C_2 \frac{dh_2}{dt} = q_1 - q_2 \quad (3.27)$$

Si q se considera la entrada y q_2 la salida, la función de transferencia es la siguiente

$$\frac{Q_2(s)}{Q(s)} = \frac{1}{R_1 C_1 R_2 C_2 s^2 + (R_1 C_1 + R_2 C_2 + R_2 C_1) s + 1} \quad (3.28)$$

Se suponen los siguientes valores para los parámetros del sistema $R_1 = 1 \text{ seg}/m^2$, $R_2 = 1.5 \text{ seg}/m^2$, $C_1 = 1 m^2$, $C_2 = 1.5 m^2$.

Las unidades de R_1 y R_2 son $\frac{m}{m^3/seg}$, lo que nos dice que por cada metro de altura en el tanque, la válvula permite que un metro cúbico de líquido por segundo salga del tanque.

3.7.1. Aplicación del filtro de Kalman al sistema de nivel de líquido.

En este caso se utilizará el filtro de Kalman como un observador de estados en presencia de ruido de medición y ruido del proceso, es decir, se considera el diagrama de bloques de la figura 3.6.

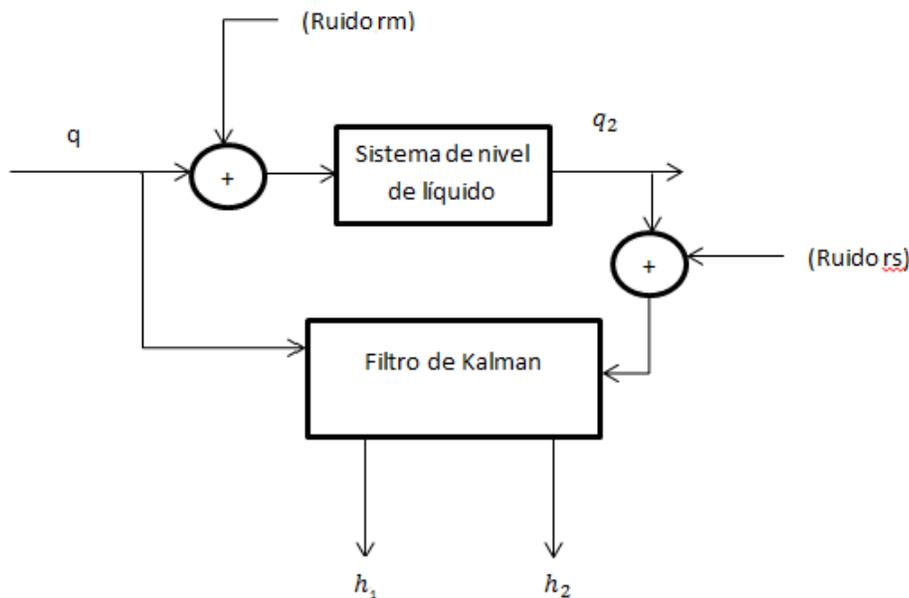


FIGURA 3.6. Diagrama a bloques del sistema de nivel de líquido

En la inicialización del código se establece un tiempo final de la prueba, un periodo de muestreo T_s , el número de muestras totales (N), se crea un vector de tiempo y se establece un valor a la covarianza del ruido de medición R , la cual posteriormente se utilizará para contaminar de ruido la señal obtenida en la medición. Después de haber hecho esto, se establecen los valores de los parámetros de la función de transferencia, se le aplica una entrada escalón unitario con ruido de varianza q_0 al sistema y se almacena la respuesta en una variable.

A continuación se describe el código desarrollado en Matlab para simular el diagrama de la figura 3.6.

- `tfinal=150;`
- `Ts=0.1;`
- `N=tfinal/Ts;`
- `t=0:Ts:tfinal-Ts;`
- `R=0.5;`
- `r1=1;`
- `r2=1.5;`
- `c1=1;`
- `c2=1.5;`
- `a=r1*c1*r2*c2;`
- `b=r1*c1+r2*c2+r2*c1;`
- `c=1;`
- `num=[1];`
- `den=[a,b,c];`
- `sys=tf(num,den);`
- `q=0.5`
- `rm=rand(1,N)*sqrt(q);`
- `u=ones(1,N)+rm;`
- `respuesta=lsim(u,t);`

Posteriormente, se contamina la señal de respuesta con un ruido de medición gaussiano, con varianza R . Se grafica la respuesta real y la contaminada para compararlas. Y se crea el vector escalón unitario sin ruido, para poder generar la entrada escalón en las ecuaciones del filtro de Kalman (ver figura 3.6)

El código, utilizado para hacer esto es el siguiente:

- `rs=randn(1,N)*sqrt(R);`
- `vs=respuesta+rs';`
- `subplot(2,1,1);`
- `plot(t,vs,'b',t,respuesta,'r');`
- `V=axis; legend('Respuesta Ruidosa','Respuesta Real');`
- `vm =ones(1,N);`

Como podemos observar en las ecuaciones del filtro de Kalman, se requiere de un modelo en el espacio de estados, este modelo puede ser obtenido a partir de las ecuaciones 3.24,3.25,3.26 y 3.27, las matrices A_c , B_c , C_c son las siguientes

$$Ac = \begin{pmatrix} -\frac{1}{R_1 C_1} & \frac{1}{R_1 C_1} \\ \frac{1}{R_1 C_2} & -\frac{1}{R_1 C_2} - \frac{1}{R_2 C_2} \end{pmatrix}, Bc = \begin{bmatrix} \frac{1}{C_1} \\ 0 \end{bmatrix} \text{ y } Cc = \begin{bmatrix} 0 & \frac{1}{R_2} \end{bmatrix}$$

El modelo 3.28 es un modelo de estado continuo, y el vector de estados es $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$ sin embargo, el algoritmo de Kalman está basado en el modelo discreto en el espacio de estados, por esta razón es necesario obtener primero la discretización del modelo 3.28 para obtener el modelo discreto

$$x_{k+1} = Ax_k + Bq_k \quad (3.29)$$

$$q_{2k} = Hx_k \quad (3.30)$$

Donde:

$$A = e^{AcT_s} \quad (3.31)$$

$$B = \left[\int_0^{T_s} e^{A\zeta} d\zeta \right] B \quad (3.32)$$

$$H = Cc \quad (3.33)$$

Las ecuaciones 3.31,3.32,3.33 representan la discretización con retenedor de orden cero, de acuerdo a Astrom and Wittenmark (1999)

El código utilizado para escribir estas matrices y pasar del modelo continuo al modelo discreto en Matlab, es el siguiente.

- $Ac = [-1/r1*c1, 1/r1*c1; 1/r1*c2, (-1/r1*c2-1/r2*c2)];$
- $Bc = [1/c1; 0];$
- $Cc = [0, 1/r2];$
- $Dc = 0;$
- $[A B C D] = c2dm(Ac, Bc, Cc, Dc, Ts, 'zoh')$

Una vez que se obtienen las matrices en el espacio de estados discreto, hay que suponer una matriz de covarianzas de los ruidos. Cabe resaltar que en la práctica, esta matriz no se conoce, por tanto aquí se supondrá con valores distintos al ruido que se metió en la medición y el ruido del proceso. La matriz de covarianza Q se supone que tiene la forma $Q = \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix}$ y la matriz de covarianza R (que en este caso es un escalar), se escriben de la siguiente forma en el código:

- $q_1=0.01; q_2=0.01;$
- $Q=[q_1 \ 0; 0 \ q_2];$
- $R=1;$

A continuación comienzan las iteraciones del filtro de Kalman, primero se inicia la matriz de covarianza del error P_k , que es una matriz diagonal, cuyo valor inicial se ajustó por prueba y error. Y el vector de estados en cero.

- $P=\text{eye}(2)*1000; x=\text{zeros}(2,N)$

Después se procede a realizar la predicción y corrección indicadas por las ecuaciones del filtro de Kalman, en el número total de muestras que se generaron con el tiempo final y el tiempo de muestreo.

Primero realizando la etapa de predicción:

$$\hat{x}_k^- = A\hat{x}_{k-1} + BV_{\text{entradaEscalón}}$$

$$P_k^- = AP_{k-1}A^T + Q$$

Una vez realizada la etapa de predicción, se precede a realizar la corrección:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \text{ (Se actualiza la ganancia de Kalman)}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (V_{\text{SalidaMedición}} - H\hat{x}_k^-) \text{ (Se actualiza el estado)}$$

$$P_k = (I - K_k H) P_k^- \text{ (Se actualiza la covarianza del error de estimación)}$$

Una vez hecho ésto, se grafican los estados, el estado x_1 , que es la altura del primer tanque y el estado x_2 que es la salida. Tomando en cuenta que la salida es $h_2 = x_2/R_2$, se puede obtener una estimación por el filtro de Kalman. De igual forma se grafica la respuesta real de h_2 mediante $\hat{h}_2 = \hat{x}_2/R_2$ donde \hat{x}_2 es el estimador de x_2 , para poder comparar los resultados y ver si se obtuvieron los resultados esperados.

El código para hacer ésto, es el siguiente:

- for k=2:N
- $x(:,k)=A*x(:,k-1)+B*vm(k-1);$
- $P=A*P*A'+Q;$
- $K=P*C'*\text{inv}(C*P*C'+R);$
- $x(:,k)=x(:,k)+K*(vs(k)-C*x(:,k));$
- $P=(\text{eye}(2)-K*C)*P;$
- end
- k=0:N-1;
- $k=k*Ts;$
- subplot(2,1,2);
- plot(k,x(1,:), 'k', k, respuesta, 'r', k, C(2)*x(2,:), 'b:');

- `axis(V)`;
- `legend('Estado 1(Llenado primer tanque)', 'Respuesta real', 'Respuesta de Kalman')`

Una vez implementado el código, podemos observar que las gráficas quedaron como aparecen en la figura 3.7.

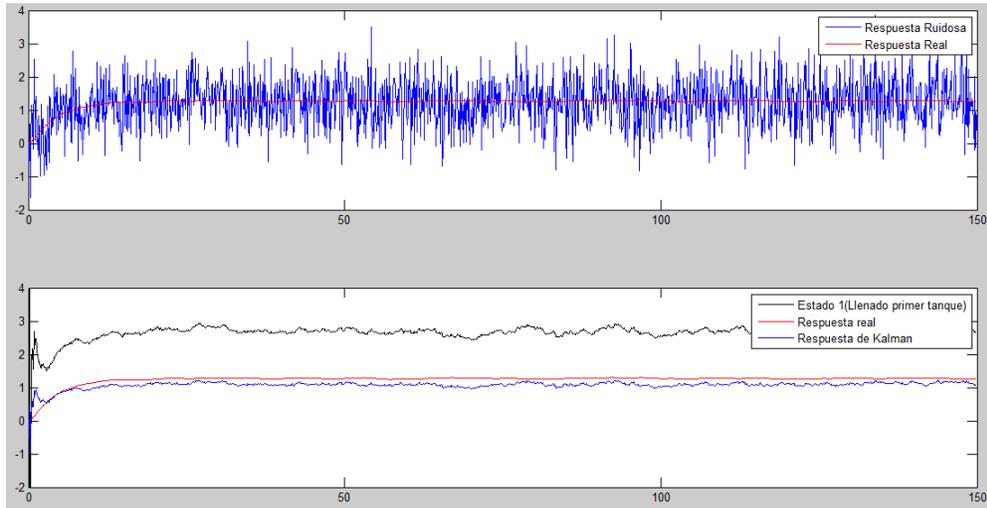


FIGURA 3.7. Resultados de la simulación

Ya que las escalas de ambas gráficas de la figura 3.7 son iguales, podemos notar que en la figura de arriba, se observa la respuesta ruidosa y la respuesta real (sin ruido). Podemos notar que existe mucho ruido, por lo cual si no se tuviera la respuesta real graficada, sería muy difícil saber que es lo que se suponía que obtendríamos en una medición para considerarla correcta. La gráfica de abajo de la figura 3.7, representan los estados estimados por el filtro de Kalman. Podemos observar el estado 1 (la altura del primer tanque) pero la parte que interesa más analizar, es la respuesta de Kalman, es decir el estado 2, ya que como se puede observar, la respuesta de Kalman, “sigue” a la respuesta real que esperábamos obtener. Si comparamos la respuesta de Kalman, con la de la medición ruidosa que aparece en la gráfica superior, notamos que el ruido ha desaparecido significativamente y se aproximó también significativamente a la respuesta real. Por tanto podemos concluir que el filtro de Kalman, es bueno para eliminar ruido, siempre y cuando conozcamos bien el modelo matemático del sistema de donde obtenemos las mediciones.

Aplicación del Filtro de Kalman a la medición de inclinación

En éste capítulo se describe una aplicación del filtro de Kalman. Se plantea la implementación de un medidor de ángulo con base en las señales recibidas de un acelerómetro y un giroscopio electrónicos. Se describe la obtención de las ecuaciones del filtro y su evaluación en simulación.

4.1. Acelerómetros

Un acelerómetro es un instrumento destinado a medir aceleraciones. Aunque si bien, por sus características es capaz de medir inclinaciones y cambios de posición en los tres ejes X, Y y Z. Se basa en el tipo de aceleración asociada al peso experimentado por una masa de prueba en el marco de referencia del dispositivo.

Actualmente se cuenta con acelerómetros digitales, que son capaces de medir estas aceleraciones, pero como siempre, existe un dispositivo antiguo del cuál se tomó la idea para la construcción del nuevo. Los acelerómetros antiguos son tan sencillos como un resorte con una bola, en donde la bola presiona al resorte y se puede medir el desplazamiento que hizo la bola basándonos en que tan “compacto” o que tan “estirado” quedó el resorte. La figura 4.1 ejemplifica ésto.

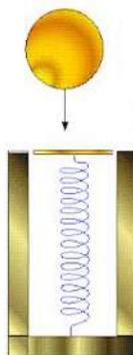


FIGURA 4.1. Resorte y Bola

Otra aplicación de éste tipo de acelerómetros, se puede en ver en un sismógrafo, en donde mediante la fuerza que ejerce la gravedad sobre la bola, comprime al resorte o lo estira y con un lápiz fijado a la bola, se puede dibujar sobre un papel si existe movimiento de la tierra que haga cambiar la posición de la bola. En la figura 4.2 se ilustra la idea descrita.

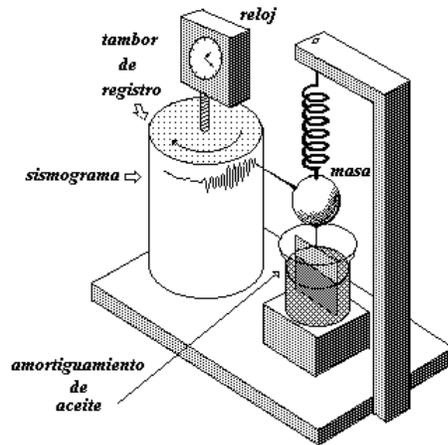


FIGURA 4.2. Sismógrafo

Ahora, con los acelerómetros electrónicos, se ha logrado encapsular todo esto en un pequeño dispositivo, el cuál es capaz de percibir cuando existe un movimiento en cualquiera de los 3 ejes (X, Y o Z) y sólo producir una señal de salida que exprese en que posición está y a donde cambió. La figura 4.3 nos da una ilustración de un acelerómetro electrónico y las direcciones en las cuales es capaz de percibir aceleraciones.

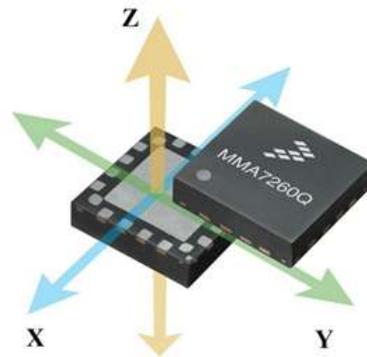


FIGURA 4.3. Acelerómetro electrónico

Actualmente existen acelerómetros electrónicos basados en diversos principios, por ejemplo: piezoeléctricos, capacitivos, de efecto hall, etc.

Como se observa en la figura 4.3, el acelerómetro es capaz de percibir aceleraciones en los tres ejes (X, Y y Z), pero también podemos colocarlo de forma en que midamos la aceleración en un sólo eje. En la figura 4.4 se observa en que forma tendría que colocarse el acelerómetro para utilizar la medición de la aceleración en un eje como una medida de inclinación.

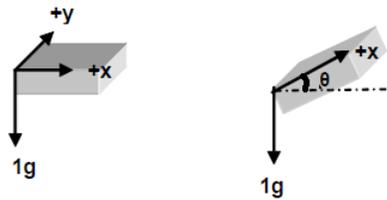


FIGURA 4.4. Acelerómetro (Un sólo eje) para medir inclinación

También se puede colocar el acelerómetro de manera que se pueda medir la inclinación en base a la medición de aceleración en dos ejes. La figura 4.5 sería un ejemplo de esto.

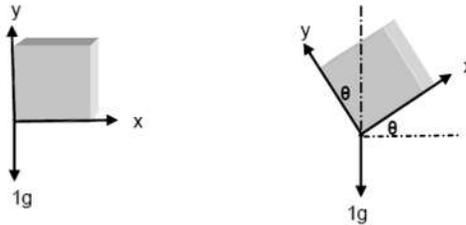


FIGURA 4.5. Acelerómetro (Dos ejes) para medir inclinación

Para la medición del ángulo de inclinación se utilizará la medición de la aceleración en un sólo eje.

Como el acelerómetro percibe la aceleración en un eje, podemos descomponer la aceleración de la gravedad (g) en componentes A_x y A_y , como se muestra en la figura 4.6. Como podemos ver las componentes en A_x y en A_y cambiarán dependiendo del ángulo θ , para un acelerómetro de un solo eje (x), la medición será

$$A_x = g \sin(\theta) \quad (4.1)$$

El acelerómetro producirá un voltaje proporcional a esta componente de aceleración.

La ecuación 4.1 expresa el voltaje ideal entregado por un acelerómetro, sin embargo, en un acelerómetro real siempre existen otras componentes, la ecuación 4.2 es una expresión típica de un acelerómetro real inclinado un ángulo θ como en la figura 4.4

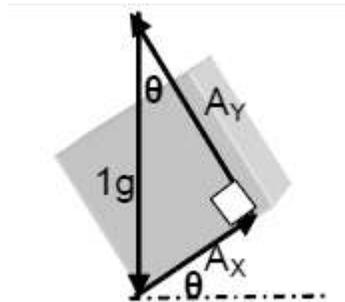


FIGURA 4.6. Descomposición vectorial

El modelo matemático de un acelerómetro es el siguiente:

$$V_{salidaA} = V_{offset} + \left(\frac{\Delta V}{\Delta g} * 1g * \sin(\theta) \right) + r_a \quad (4.2)$$

Donde:

$V_{salidaA}$ = Salida del acelerómetro en Volts

V_{offset} = Offset del acelerómetro $0g$

$\frac{\Delta V}{\Delta g}$ = Sensitividad

$1g$ = Gravedad de la tierra

θ = Ángulo de inclinación

r_a = Ruido en la señal de acelerómetro

Resolviendo para el ángulo:

$$\theta = \arcsin\left(\frac{V_{salidaA} - V_{offset}}{\frac{\Delta V}{\Delta g}}\right) \quad (4.3)$$

Ahora bien, si en 4.2 se considera $V_{offset} = 0$ y $\frac{\Delta V}{\Delta g} = 1/g$, podemos reescribir la ecuación de la siguiente manera:

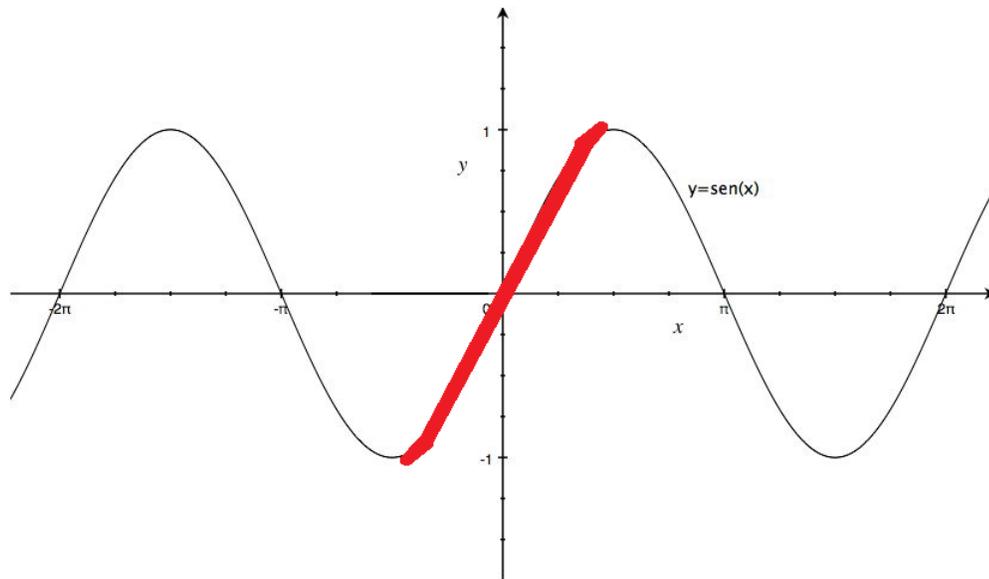


FIGURA 4.7. Gráfica de la aproximación del seno

$$V_{\text{salida}A} = \sin(\theta) + r_a \quad (4.4)$$

Notamos que 4.4 es una función no lineal de θ , sin embargo, existe un intervalo de valores de θ , donde se puede aproximar a una función lineal. En la figura 4.7 observamos la gráfica de la función $\text{sen}(\theta)$ y como en la parte en que la señal se ve un poco mas gruesa se comporta de una manera aproximadamente lineal.

Si consideramos, de la figura 4.7, que el seno se puede tratar como lineal en un intervalo de su gráfica, podemos reescribir la ecuación de la siguiente manera para $-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}$:

$$V_{\text{salida}A} \approx \theta + r_a \quad (4.5)$$

Donde podemos notar que hemos eliminado el elemento de la ecuación que era no lineal, por una proximación en un intervalo solamente.

Para nuestra simulación 4.5 será suficiente. Si llegáramos a querer mas precisión se tendría que utilizar la función $\sin(\theta)$ en lugar de θ en radianes.

4.2. Giroscopio

Un giroscopio nos sirve para medir, la orientación en el espacio de algún aparato o vehículo. Está formado principalmente por un cuerpo que gira alrededor de un eje de simetría.



FIGURA 4.8. Ejemplo de un giroscopio

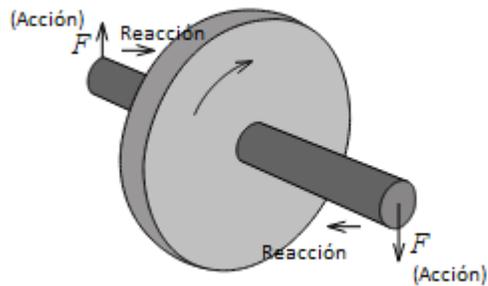


FIGURA 4.9. Rueda girando sobre un eje mostrando las fuerzas de reacción por efecto giroscopio

Una manera sencilla de describir el efecto giroscopio es con una rueda como la mostrada en la figura 4.9, la cual gira a una velocidad constante sobre un eje fijo. Si el observador trata de hacer girar el eje hacia la derecha (bajando la mano derecha y subiendo la mano izquierda) sentirá un comportamiento muy curioso, ya que el giróscopo empuja su mano derecha y tira de su mano izquierda. El observador acaba de sentir el efecto giroscópico.

El modelo matemático de un giroscopio ideal se puede escribir de la siguiente forma:

$$V_{salidaG} = \omega + b + r_g \quad (4.6)$$

Donde:

$V_{salidaG}$ = Voltaje de salida del giroscopio

ω = Velocidad angular

b = Offset del giroscopio

r_g = Ruido en la señal giroscopio

4.3. Aplicación del filtro de kalman

Ahora que tenemos una idea de lo que es un acelerómetro y un giroscopio, se simulará un sensor de ángulo de inclinación basándonos en éstos dos dispositivos. Es importante suponer que los ruidos de ambos componentes (giroscopio y acelerómetro) tienen la característica de ser ruidos blancos.

Se hará el filtrado de una secuencia de mediciones ruidosas que se obtendrían de un acelerómetro y de un giroscopio de un objeto que sufre una inclinación en un eje. Se realizará el filtrado basándonos en un filtro de Kalman.

4.3.0.1. Modelo de estado continuo. Tomando en cuenta 4.5 y 4.6 y recordando que $\omega = \frac{d\theta}{dt}$ se puede escribir una ecuación de estado continuo que relaciona los voltajes medidos con la variable de interés θ :

$$\frac{d\theta}{dt} = V_{salidaG} - b - r_g \quad (4.7)$$

y la correspondiente ecuación de salida será

$$y(t) = V_{salidaA} = \theta(t) + r_a \quad (4.8)$$

por otro lado, el offset del giroscopio (b) es desconocido, pero podemos considerar que permanece constante, es decir,

$$\frac{db}{dt} = 0 \quad (4.9)$$

con lo cual tenemos otra ecuación de estado.

4.3.0.2. Modelo de estado discreto. Si se tiene un periodo de muestreo T_s , podemos escribir el modelo de estado discreto basándonos en aproximar la integral $\int \omega(t) dx$ mediante rectángulos de ancho T_s , entonces

$$\theta(k) = \int_0^k \omega(t) dt \approx \sum_{i=0}^k \omega_i T_s \quad (4.10)$$

es decir, en forma recursiva $\theta_k = \theta_{k-1} + \omega_k T_s$, por lo tanto (tomando en cuenta 4.7)

$$\theta_k = \theta_{k-1} - b_k * T_s + V_{salidaG_k} * T_s + r_{g1k} \quad (4.11)$$

y de 4.9, obtenemos

$$b_k = b_{k-1} + r_{g2k} \quad (4.12)$$

Donde:

r_{g1k} =Es la porción del ruido r_g que afecta a θ

r_{g2k} =Es la porción del ruido r_g que afecta a b

Por lo tanto la ecuación de salida sería

$$y_k = V_{salidaA} = \theta_k + r_{ak} \quad (4.13)$$

También podemos decir entonces que las matrices de covarianza son

$$R = r_a \quad (4.14)$$

$$Q = \begin{bmatrix} r_{g1k} & 0 \\ 0 & r_{g2k} \end{bmatrix} \quad (4.15)$$

4.3.1. Simulación en Matlab. Para realizar la simulación en MATLAB primero es necesario definir un intervalo de tiempo de simulación, un periodo de muestreo, generar el número total de muestras, crear un vector de tiempo que concuerde con nuestro periodo de muestreo y definir valores para las covarianzas R y Q . El siguiente es el código para realizar esto:

- tf=200;
- Ts=0.008;
- N=tf/Ts;
- t=0:Ts:tf-Ts;

Ahora se generarán las señales de los dos sensores, para la señal del acelerómetro se considerará una velocidad de cambio de la inclinación que llamaremos ω y estará dada como $\omega = 2\pi * 0.01$, el 0.01, la variación del ángulo considerada es $\theta = \sin(\omega t)$ también se considera que el voltaje entregado por el acelerómetro está contaminado por ruido gaussiano de varianza r_a , es decir, $V_{salidaA} = \theta + r_a$.

Con esto se puede graficar el voltaje del acelerometro contaminado por el ruido, así como el ángulo θ real. El siguiente código realiza esto:

- `theta = sin(omega*t);`
- `ra=randn(1,N)*sqrt(R);`
- `va=theta+ra;`
- `subplot(2,1,1);`
- `plot(t,va,'b',t,theta,'r');`
- `legend('\theta_{acc}','\theta_{real}');`

Después se genera la señal del giroscopio, donde se considera una varianza $q = 0.3$ para el ruido r_g que con contamina a la señal de salida del giroscopio además de, un offset (también llamado bias) constante entonces

$$V_{salidaG} = \omega \cos(\omega t) + b + r_g$$

También se escriben las matrices del modelo de estado, necesarias para el filtro de Kalman A, B, H . Donde

$$A = \begin{pmatrix} 1 & -T_s \\ 0 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} T_s & 0 \end{pmatrix}$$

$$H = \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

Y la matrices de covarianza de los ruidos serían

$$Q = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}$$

$$R = 0.5.$$

R se supone constante. El código para hacer todo esto, es el siguiente:

- `q=0.3; bias=1;`
- `vg = omega*cos(omega*t)+bias+randn(1,N)*sqrt(q);`
- `A=[1 -Ts;0 1];`
- `B=[Ts 0]';`
- `H=[1 0];`
- `q1=0.1; q2=0.1;`
- `Q=[q1 0;0 q2];`
- `R=1;`

A continuación se inician las iteraciones del algoritmo del filtro de Kalman. Se inicializa una matriz de covarianza del error $P_k = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$ y un vector x_k de ceros para los estados.

Después se procede a realizar la predicción y corrección indicadas por las ecuaciones del filtro de Kalman, en el número total de muestras que se generaron con el tiempo final y el tiempo de muestreo.

Primero realizando la etapa de predicción:

$$\hat{x}_k^- = A\hat{x}_{k-1} + BV_{salidaG_{k-1}}$$

$$P_k^- = AP_{k-1}A^T + Q$$

Una vez realizada la etapa de predicción, se procede a realizar la corrección:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \text{ (Se actualiza la ganancia de Kalman)}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (V_{salidaA_k} - H\hat{x}_k^-) \text{ (Se actualiza el estado)}$$

$$P_k = (I - K_k H) P_k^- \text{ (Se actualiza la covarianza del error de estimación)}$$

Una vez realizados los cálculos de las ecuaciones del filtro de Kalman, graficamos lo obtenido para θ estimada de Kalman, la θ real y el offset estimado. El código para realizar esto, es el siguiente:

- P=eye(2)*100;
- x=zeros(2,N);
- for k=2:N
- x(:,k)=A*x(:,k-1)+B*vg(k-1);
- P=A*P*A'+Q;
- K=P*C'*inv(C*P*C'+R);
- x(:,k)=x(:,k)+K*(va(k)-x(1,k));
- P=(eye(2)-K*C)*P;
- end
- k=0:N-1;
- k=k*Ts;
- subplot(2,1,2);
- plot(k,x(1,:),'b',k,theta,'r',k,x(2,:),'k:');
- legend('\theta_{Kalman}', '\theta_{real}', 'b_{estim}')

En la figura 4.10 se observa el resultado de la simulación.

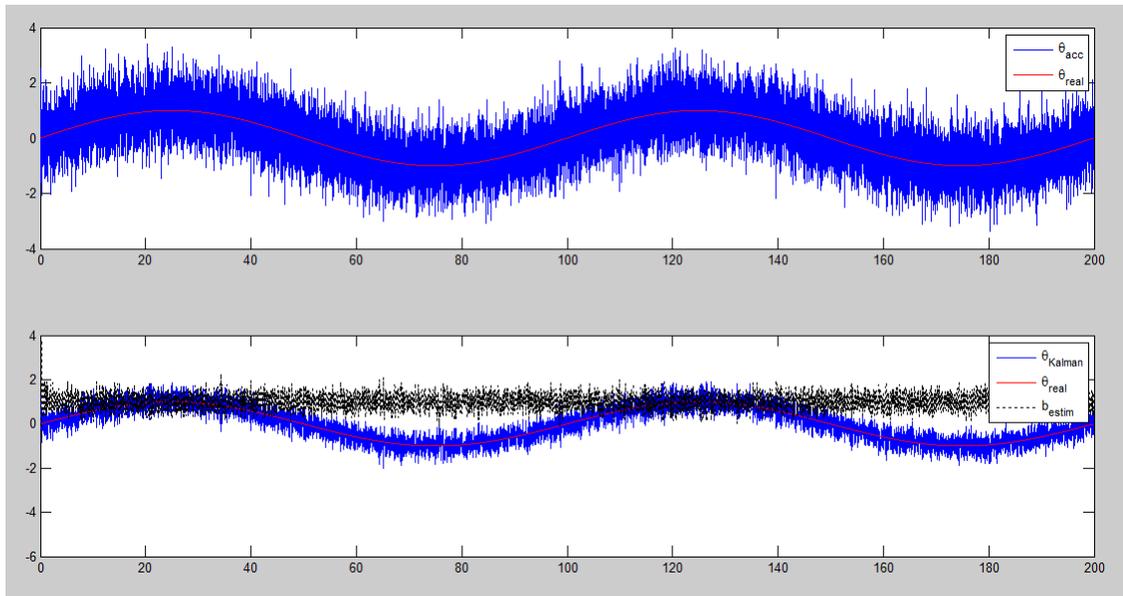


FIGURA 4.10. Resultado de la simulación

En la figura 4.10 se observa en la parte de arriba las señales simuladas que se reciben del acelerómetro y del ángulo θ , es esta gráfica se observa la gran amplitud del ruido de medición. En la parte de abajo se ve la estimación del offset, el cual tiene una tendencia constante como se había previsto. Además en la gráfica inferior se puede apreciar que la estimación del ángulo θ (θ_{Kalman}) ha disminuido en gran medida el efecto del ruido contaminante, obteniéndose una señal mucho más limpia aunque no completamente libre de ruido.

4.4. Conclusiones

Podemos observar que el objetivo de la tesis se logró, ya que a partir de conocer el modelo de la señal que queremos medir, aún cuando ésta estaba contaminada por ruido, fuimos capaces de filtrar la señal de tal manera que se eliminara parte del ruido que no nos permitía ver con claridad que era lo que se esperaba de la señal. Creo que es muy provechoso haber logrado esto, debido a que el ruido que contamina a una señal es algo de lo cuál no podemos huir los que trabajamos con señales analógicas o digitales. Y gracias a este filtro, si se conoce el modelo que describe al proceso, podemos eliminar el ruido. También pudimos observar que las estimaciones basadas en probabilidad son buenas, debido a que el filtro funciona de manera adecuada.

Conclusiones

5.1. Conclusiones Generales

En este trabajo que se llevó a cabo sobre el filtro de Kalman y su aplicación a la medición de inclinación mediante un acelerómetro y un giroscopio electrónicos, se considerará un objetivo principal, que es probar el desempeño del filtro de Kalman, y varios secundarios. Para los objetivos secundarios, se logró realizar un resumen comprensible con ejemplos concisos que permitirán a un estudiante de ingeniería ser capaz de entender la teoría de la probabilidad necesaria para el filtro de Kalman.

Se presentó el algoritmo de Kalman donde se explicaron las ecuaciones que lo gobiernan y se hicieron diagramas de bloques para que quedara más comprensible su funcionamiento.

Finalmente se evaluaron simulaciones de un sistema de nivel de líquido y de la medición de inclinación mediante un acelerómetro y un giroscopio electrónicos, de los cuales se obtuvieron resultados favorables, es decir, se obtuvo un buen desempeño del filtro de Kalman, con lo cual cumplió el objetivo principal de la tesis.

Debido a las simulaciones realizadas, podemos observar que el filtro de Kalman es bueno para reducir de manera significativa el ruido que puede existir en un proceso, en la medición de su salida o en ambos.

Esta tesis puede tener un gran impacto para la eliminación del ruido en la medición y ruido en el proceso, si se sigue desarrollando proyectos tomándola como base, se podría lograr que en cualquier proceso en el cual estemos trabajando, podamos disminuir el ruido que exista, de tal manera que la respuesta que obtenemos de él, sea la más confiable posible. Un ejemplo de esto, es el acelerómetro y el giroscopio que se utilizaron en esta tesis, ya que tienen por naturaleza una respuesta “ruidosa”, por lo tanto, con un modelo del proceso, se fue capaz de eliminar el ruido de manera significativa.

Es importante no olvidar una de las grandes limitaciones del filtro de Kalman, la cual consiste en que requiere el modelo de espacio de estado del proceso del

cual se quiere eliminar el ruido. Esta limitación puede ser un gran impedimento para aplicarlo si no conocemos el modelo del proceso. Otra limitante es que el modelo debe ser lineal, sin embargo esta limitante puede ser superada aplicando la versión extendida del filtro de Kalman, la cual no fue tratada en esta tesis.

Trabajo futuro.

Un trabajo interesante sería pasar de la etapa de simulación al código en tiempo real y la implementación del algoritmo del filtro de Kalman en un microcontrolador para aplicar la estrategia a un par acelerómetro-giroscopio de manera real.

Otro trabajo interesante es pasar a la versión extendida del filtro de Kalman para poder tratar el par acelerómetro-giroscopio en su región no lineal.

A partir del presente trabajo también se sientan los precedentes para incursionar en otras aplicaciones del filtro de Kalman, tales como el filtrado de señales de un sistema GPS, diseño y construcción de una unidad de medición inercial (IMU) para aviones o helicópteros experimentales, etc.

Bibliografía

- Astrom, K. J. and Wittenmark, B. (1999). *Computer-Controlled systems: theory and design*. Prentice hall.
- Bishop, G. and Welch, G. (2001). An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8:27599–3175.
- Brown, R. G., Hwang, P. Y., et al. (1992). *Introduction to random signals and applied Kalman filtering*, volume 1. John Wiley & Sons New York.
- Ham, F. M. and Brown, R. G. (1983). Observability, eigenvalues, and kalman filtering. *Aerospace and Electronic Systems, IEEE Transactions on*, (2):269–273.
- Jacobs, O. L. R. and Jacobs, O. (1993). *Introduction to control theory*, volume 2. Oxford University Press Oxford.
- Kalman, R. E. (1960). *A new approach to linear filtering and prediction problems*. Transactions of the ASME-Journal of Basic Engineering, pp. 35-45.
- Ogata, K. (2003). *Ingeniería de control moderna*. Pearson Educación.
- Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables and stochastic processes*. Tata McGraw-Hill Education.
- Peter, S. M. (1979). *Stochastic models, estimation, and control*. vol. 1.