EL2320 - Introduction to Estimation from Measurement

John Folkesson

Introduction to Estimation (Chap 1-2 in Thrun)

KTH
Centre for Autonomous Systems

What is Estimation

- Get an imperfect numerical value to describe of some part of the world.
- We actually estimate all the time:
  - time
  - temperature
  - distance
- Here we will generate estimates based on sensor measurements.
- Dealing with the uncertainty (noise) “correctly” is the focus of this course.

Estimate Uncertainty

- We often include estimates of the uncertainty along with the number we estimate: bounds, confidence measure, variance ...
- Sources of uncertainty:
- measurement error, sample variation, disturbances, noise.
- algorithmic error, that is error in how we estimate or model the problem.
- There can even be errors in our estimates of our uncertainty, or error in our estimated error.

Statistical Inference aka Induction

- Inference = Draw logical conclusions.
- Statistical = Numerical data (typically containing random variations).
- Try to understand the process and the estimation errors.
- Build a 'Stocastic Model' of the process.
- Build an 'Estimator' based on the model.
- Hopefully: Prove the estimator is unbiased and optimal.

Let X be a discrete random variable:

- X can only take discrete values (e.g. heads or tails)
- $P(X = x)$ denotes the probability of X being x
- $0 \leq P(X = x) \leq 1$.
- $\sum_x P(X = x) = 1$
- often write $P(X = x)$ as simply $P(x)$: read Probability of x.
- $P(x|z)$: read Probability of x given z.
  Which actually means: probability that random variable X is equal to x if we know for certain that random varible Z is equal to z.

Let X be a continuous random variable:

- X can take any value in $[a, b]$ often $(-\infty, \infty)$
- (Cumulative) Distribution function: $F(x) = P(X \leq x)$
- $F(a) = 0$
- $F(b) = 1$
- Probability density function (pdf): $p(x) = \frac{dF(x)}{dx}$
- $\int_v^u p(x)dx = P(v \leq X \leq u) = F(u) - F(v)$
- $\int_a^b p(x)dx = 1;$

- Expectation value of X:

  $E[X] = \bar{x} = \sum_x xP(X = x)$

  Or for continuous pdf's: $E[X] = \int_{-\infty}^{\infty} xp(x)dx$

- For scalar X variance:

  $\sigma^2 = E[(x - \bar{x})^2]$

  $\sigma^2 = \int_{-\infty}^{\infty} (x - \int_{-\infty}^{\infty} x'p(x')dx')^2 p(x)dx$

  $\sigma^2 = E[x^2] - \bar{x}^2$ (show this)

- The variance is nonnegative

Write answers

We read 20°C on a calibrated digital thermometer displaying to degrees while its actual temperature sensor is accurate to hundredths of a degree. If the true temperature is denoted by T then what is:

1. $P(T \leq 20.0 | read20) =?$
2. $P(T \geq 19.5 | read20) =?$
3. $P(T \leq 19.75 | read20) =?$
4. $P(T = 20.0 | read20) =?$
5. $E(T | read20) =?$
6. variance of T?

- Write out answers as best as you can. (3 min.)

Let X and Y be two random variables

- Joint distribution function:
  $F(x, y) = P((X \leq x) \cap (Y \leq y))$
- Joint pdf: $p(x, y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$
- $F(x) = F(x, \infty)$
- Marginalization: $p(x) = \int_{-\infty}^{\infty} p(x, y) dy$
- Marginalization: $p(x) = \int_{-\infty}^{\infty} p(x, y) dy$

- Conditional probability:
  $p(x|y)p(y) = p(x,y)$
- $p(x|y) = \frac{p(x,y)}{p(y)}$ so long as $p(y) > 0$
- what is $p(x|y)$ when $p(y) = 0$?
- Theorem of total probability:
  - Discrete: $P(x) = \sum_y P(x|y)P(y)$
  - Continous: $p(x) = \int_{-\infty}^{\infty} p(x|y)p(y)dy$

P(X,Y)	X=1	X=2
Y=1	.1	.5
Y=2	.3	.1

What is $P(X = 1)$? (hint marginalize out the Y)
What is $P(Y = 1|X = 2)$?

- $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$
- Bayes rule:
  - Discrete: $P(x|y) = \frac{P(y|x)P(x)}{P(y)}$
  - Continous: $p(x|y) = \frac{p(y|x)p(x)}{p(y)}$
- so long as $p(y) > 0$
- we say x and y are independent of one another if: $p(x|y) = p(x) \Leftrightarrow p(x, y) = p(y)p(x) \Leftrightarrow p(y|x) = p(y)$

P(X,Y)	X=1	X=2
Y=1	.1	.5
Y=2	.3	.1

Are X and Y independent?

P(X,Y)	X=1	X=2
Y=1	$\frac{3}{30}$	$\frac{7}{30}$
Y=2	$\frac{6}{30}$	$\frac{14}{30}$

Are X and Y independent?

- $C_{xy} = E[(x - \bar{x})(y - \bar{y})]$ is called the covariance of X and Y.
- $E[xy] - \bar{x}\bar{y}$ is related to the correlation coefficient of X and Y (divide by $\sigma_X \sigma_y$.
- X and Y said to be uncorrelated if
  $< E[xy] = \bar{x}\bar{y}$.
  which implies $C_{xy} = 0$.
- For vectors x covariance matrix:
  $\Sigma = E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T]$
  is positive semidefinate, (Eigenvalues $\lambda \geq 0$).

$C_{xy} = E[(x - \bar{x})(y - \bar{y})]$ is called the covariance of X and Y.

P(X,Y)	X=1	X=2
Y=1	.1	.5
Y=2	.3	.1

What is the covariance of X and Y?
Are they correlated?

Law of Large Numbers

The average of a series of n unbiased measurements of a value x

$\mu_n = \frac{1}{n} \sum_1^n z_n$ will approach x as.

$$\lim_{n \to \infty} \mu_n = x$$

μ_n is a random variable so we need to be a bit careful with the limit. There are two ways to do limits of random variables leading to the Weak and strong versions of the law.

Weak: $\lim_{n \to \infty} P(|\mu_n - x| < \epsilon) = 1$

Strong: $P(\lim_{n \to \infty} \mu_n = x) = 1$

Carl Friedrich Gauss invented the normal distribution in 1809 to help explain the method of least squares.

The scalar X is a Gaussian or normal variable if its pdf is of the form:

$p(x) = G(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

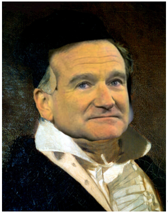
(Laplace figured out the $\frac{1}{\sqrt{2\pi\sigma^2}}$)

where the mean of X is

$\mu = E[x] = \bar{x}$

and the variance of X is $\sigma^2 = E[(x - \bar{x})^2]$

We say X is $N(\mu, \sigma^2)$ or $x \sim N(\mu, \sigma^2)$

Carl Friedrich Gauss invented the normal distribution in 1809 to help explain the method of least squares.

The scalar X is a Gaussian or normal variable if its pdf is of the form:

$$p(x) = G(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(Laplace figured out the $\frac{1}{\sqrt{2\pi\sigma^2}}$)

where the mean of X is

$$\mu = E[x] = \bar{x}$$

and the variance of X is $\sigma^2 = E[(x - \bar{x})^2]$

We say X is $N(\mu, \sigma^2)$ or $x \sim N(\mu, \sigma^2)$

Central Limit Theorem

Let $z_1, z_2, \ldots, z_n$ be a sequence of n independent and identically distributed random variables.

Let x be the distribution's mean and $q > 0$ its variance.

The central limit theorem states that as the sample size n increases the distribution of the sample average of these random variables approaches the normal distribution with a mean x and variance q/n irrespective of the shape of the common distribution.

Our first estimation algortihm:

$\bar{z}_n = \frac{\sum z_i}{n} = \frac{\bar{z}_{n-1}(n-1)+z_n}{n}$

$\sigma_n^2 = q/n = \sigma_{n-1}^2(n-1)/n.$

a SICK example.