Bayesian Approach

Probability function of event A in sample space S:

Kolmogorov's Axioms: $P(A) \geq 0 \ \forall A$, $P(S) = 1$,

$P(\bigcup_i A_i) = \sum_i P(A_i)$ if A_i are disjoint (mutually exclusive)

Joint probability: discrete example for events A and B

Total 100 observations made.

Outcome logged in a contingency table.

	A	not A
B	55	5
not B	10	30

$$p(A) = \frac{55+10}{100} = 0.65 \qquad p(!A) = 0.35$$

$$p(B) = 0.6 \qquad p(!B) = 0.4$$

Conditional probability:

$$p(A|B) = \frac{55}{55+5} = 0.917$$

probability of A given that event B occurs.

$$p(A \cap B) = \frac{55}{100}$$

probability that both A, B occurs.

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{0.55}{0.6} = 0.917$$

Bayes' Theorem:

$$p(A \cap B) = p(A|B)\,p(B) = p(B|A)\,p(A).$$

$$\therefore \quad p(B|A) = \frac{p(A|B)\,p(B)}{p(A)}$$

If A and B are not independent, i.e. they are correlated. Given information from observing event B, one can infer about event A.

Note: two events A and B are statistically independent if $p(A \wedge B) = p(A)\, p(B)$

$\rightarrow p(A|B) = p(A)$ and $p(B|A) = p(B)$

Example: A = drug test positive
B = athlete used drug

prior knowledge: estimate 60% of athletes do drugs.
$$P(B) = 0.6$$

Drug test result bring extra information:

$p(A|B) = 0.917$ 91.7% test positive for actual drug users

$p(A|\,!B) = 0.25$ 25% False alarm rate

Since A and B are correlated (not independent) a "positive drug test" increases the probability of "athlete used drug"

$$\underbrace{p(B|A)}_{\text{posterior}} = \underbrace{\frac{p(A|B)}{p(A)}}_{\text{"evidence" or "support" for B given A}} \underbrace{p(B)}_{\text{prior}} = \frac{0.917}{0.65}\, 0.6 = 0.846$$

Note: $p(A|B)$ is also the likelihood function $\mathcal{L}(B|A)$

$p(A) = p(A|B)\,p(B) + p(A|\,!B)\,p(!B)$ is a normalizing term.

Conclusion: the athlete is found guilty more easily given his positive drug test result.

Multivariate, Continuous case:

probability density function (pdf) of a Gaussian random variable $X \sim N(x^b, B)$

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |B|}} \exp\left(-\frac{1}{2}(x^b - x)^T B^{-1} (x^b - x)\right)$$

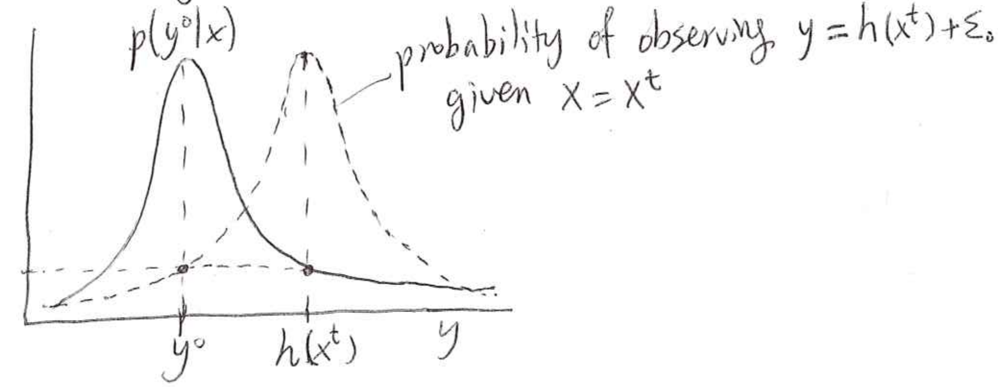
$p(x)$ is the probability of $x = x^t$, prior distribution

x^b is prior mean (first moment)

B is prior error covariance (second moment)

Note: n-th moment $= \int (x - x^b)^n p(x)\, dx$, for $n > 1$

$p(y^o|x)$ is the probability of observing a value y^o given that $x = x^t$. It is also the likelihood function of $x = x^t$ given that y^o is observed. $\mathcal{L}(x|y^o)$



$$p(y^o|x) = \frac{1}{\sqrt{(2\pi)^p |R|}} \exp\left(-\frac{1}{2}(y^o - h(x))^T R^{-1} (y^o - h(x))\right)$$

"observation likelihood"

According to Bayes' Theorem, posterior distribution

$$p(x|y^o) = \frac{p(y^o|x)\, p(x)}{\int p(y^o|x)\, p(x)\, dx} \quad = p(y^o) \text{ is a normalizing factor.}$$

multiplication of two Gaussian pdf → a Gaussian pdf

$p(x|y^o) = N(x^a, A)$ can show that:

$$x^a = x^b + W(y^o - h(x^b))$$

$$A = (I - WH)B$$

$$W = BH^T(HBH^T + R)^{-1}$$

Proof in one-variable case:

$$-2\ln(p(x|x^o)) = -2\ln(p(x^o|x)\,p(x)) + C_1$$

$$\frac{(x-x^a)^2}{\sigma_a^2} = \frac{(x-x^b)^2}{\sigma_b^2} + \frac{(x^o-x)^2}{\sigma_o^2} + C_1$$

$$= \frac{x^2 - 2xx^b + x^{b2}}{\sigma_b^2} + \frac{x^2 - 2xx^o + x^{o2}}{\sigma_o^2} + C_1$$

$$= \left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right)x^2 - 2x\left(\frac{x^b}{\sigma_b^2} + \frac{x^o}{\sigma_o^2}\right) + C_2$$

complete the square

$$= \underbrace{\left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}\right)}_{=\frac{1}{\sigma_a^2}}\left(x^2 - 2x\,\underbrace{\frac{\frac{x^b}{\sigma_b^2} + \frac{x^o}{\sigma_o^2}}{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}}}_{= x^a = \frac{x^b\sigma_o^2 + x^o\sigma_b^2}{\sigma_b^2 + \sigma_o^2}} + x^{a2}\right) + C_3$$