Ex6-1.

Bayes' theorem: $p(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(A,B)}{\int P(B|A)p(}$

probability: $\theta \rightarrow p(A)$

likelihood: $p(A|\theta)$

$p(B) = \sum_j p(B|A_j)\,p(A_j)$

$p(A_i|B) = \frac{p(B|A_i)\,p(A_i)}{\sum_j p(B|A_j)\,p(A_j)}$

$\uparrow$

θ: continus.

likelihood

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{\int p(D|\theta)\,p(\theta)\,d\theta} = \frac{p(D,\theta)}{\int p(D,\theta)\,d\theta}.$$

$\leftarrow$ Prior

$\uparrow$

posterior

$\leftarrow$ normalization

(a) $KL(Q(\theta)||p(\theta|D)) + L$

$$= \int Q(\theta) \log \frac{Q(\theta)}{p(\theta|D)}\,d\theta + \int Q(\theta) \log \frac{P(\theta,D)}{Q(\theta)}\,d\theta.$$

$$= \int Q(\theta) \cdot \log \frac{Q(\theta)}{p(\theta|D)} \cdot \frac{p(\theta|D)\,p(D)}{Q(\theta)}\,d\theta.$$

$$= \int Q(\theta) \cdot \log p(D)\,d\theta.$$

$$= \log p(D) \int Q(\theta)\,d\theta = \log p(D)$$

(b) lower bound:

Let $(A, \leq)$ is a partial ordered set.

z is a lower bound if $\exists z \in A$.

s.t. $\forall x \in B \sqsubseteq A$, $x \geq z$.

$$KL(Q(\theta) \| P(\theta|D)) = \int Q(\theta) \cdot \log \frac{Q(\theta)}{P(\theta|D)} d\theta$$

Gibbs' inequality:

$P = \{p_1, \cdots, p_n\}$

$Q = \{q_1, \cdots, q_n\}$

$$-\sum_{i=1}^{n} p_i \log_2 p_i \leq -\sum_{i=1}^{n} p_i \log_2 q_i$$

$$KL(Q(\theta) \| P(\theta|D)) \geq 0$$

$$KL + L = \log p(D)$$

$L = \log p(D) - KL \leq \log p(D) \Rightarrow L$ is a lower bound.

with $KL(P \| Q)$ zero iif $P = Q$ almost everywhere.

a.e.

Ex 6-2

(a) $$p(D|\theta) = \prod_{i=1}^{n} p(d_i | \theta, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(d_i-\theta)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(d_i-\theta)^2\right\}$$

(b) $$p(\theta|D) = N(\mu_m, \sigma_m^2)$$

$$\mu_m = \frac{\sigma^2}{n\sigma_p^2+\sigma^2}\mu_p + \frac{n\sigma_p^2}{n\sigma_p^2+\sigma^2}\left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)$$

$$\frac{1}{\sigma_m^2} = \frac{1}{\sigma_p^2} + \frac{n}{\sigma^2}.$$

$\mu_p = 0$, $\sigma_p = 1$. $\Rightarrow$ $$\frac{1}{\sigma_m^2} = \frac{1}{1} + \frac{n}{1} = n+1$$

$\sigma^2 = 1$

$$\sigma_m^2 = \frac{1}{n+1}$$

$$\mu_m = \frac{1}{1+1}\cdot 0 + \frac{n}{n+1}\left(\frac{1}{n}\sum_{i=1}^{n} d_i\right) = \frac{1}{n+1}\sum_{i=1}^{n} d_i$$

$$p(\theta|D) = N\left(\frac{1}{n+1}\sum_{i=1}^{n} d_i, \frac{1}{n+1}\right)$$

(c)

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{p(D)}$$

$$= \frac{\prod_{i=1}^{n} N(\theta,\sigma^2)\cdot N(\mu_p,\sigma_p^2)}{\int \prod_{i=1}^{n} N(\theta,\sigma^2)\cdot N(\mu_p,\sigma_p^2)\,d\theta}.$$

If $p(\theta)$ changes, then $p(\theta)$'s conjugated prior might not have a analytical solution.

Gaussian likelihood is self conjugated.

conjugated distribution: prior dist ~~as~~ same as posterior dis

conjugated priors leads to closed-form expression for the posterior.