Problem 1

a) We shall assume that $c\,q(x) \geq \tilde{p}(x)$, $\forall x$

Until termination (e.g exhausted comp. res.):

1) Draw $x_0 \sim q(x)$

2) Draw $u_0 \sim Uni(0, kq(x_0))$

3) If $u_0 > \tilde{p}(x_0)$ then reject the sample, accept otherwise with prob:

$$p(\text{accept}) = \frac{1}{k} \int \tilde{p}(x)\, dx$$

b) Yes, they are indep (Ian Murray, 2004). Steps are indep of prev. samples.

c) $$E[f] = \int f(z)\, p(z)\, dz = \int f(z) \frac{p(z)}{q(z)} q(z)\, dz \approx \frac{1}{L} \sum_{i=1}^{L} \frac{p(x_i)}{q(x_i)} f(x_i) =$$

$$= \frac{1}{L} \sum_{i=1}^{L} w_i f(x_i), \quad \text{where } w_i = \frac{p(x_i)}{q(x_i)}$$

d) $$\alpha(x_{t+1}, x_t) = \min\left(1, \frac{\tilde{p}(x_{t+1})\, q(x_t)}{\tilde{p}(x_t)\, q(x_{t+1})}\right) = \min\left(1, \frac{p(x_{t+1}) \cancel{Z_p} q(x_t)}{p(x_t) \cancel{Z_p} q(x_{t+1})}\right) =$$

$$= \min\left(1, \frac{p(x_{t+1})\, q(x_t)}{p(x_t)\, q(x_{t+1})}\right)$$

f) By rejecting x_2 and x_5 we leave them as prev. samples for the next iteration, thuss, the sequence will be: x_1, x_1, x_3, x_4, x_4

g) 1) Rejection sampler → Does not scale well, because as dim grows, we have a large space (volume) between $\tilde{p}$ and $\tilde{q}$, thus many samples will be rejected.

2) Importance sampler → Does not scale well because as # of dimensions increases it becomes hard to find $q(x)$ that fits to $p(x)$, and thus a poor choice will make it hard to generate samples such that they are both probable according to $p(x)$ and $q(x)$.

as $w_i = \frac{p(x_i)}{q(x_i)}$ $\quad E[f] \approx \frac{\sum_i w_i f(x_i)}{\sum_j w_j}$,

3) Independance sampler → Scale much better because Markov perspective leads to efficient decompositions of high-dimensional problems in a sequence of smaller problems that are much easier to solve.

e) Even though we get samples $x \sim q(x)$ which do not depend on the prev. samples our threshold α still takes into account prev. samples, and thus the decision becomes dependent on the pevious draws. Therefore, generated samples are not independent.

Problem 5

a) $E[\mathbf{x}] = \begin{pmatrix} E[x_1] \\ \vdots \\ E[x_n] \end{pmatrix}$, since $x_i \sim \text{Bern}(\mu_i)$, $E[x_i] = 1\mu_i + 0(1-\mu_i) = \mu_i$

thus $E[\vec{x}] = \vec{\mu}$

b) Since Bernoulli trials are independent, we'll get a diagonal matrix with variances along diagonal.

So $\text{Var}[x_i] = E[x_i^2] - (E x_i)^2 = \mu_i - \mu_i^2 = \mu_i(1-\mu_i)$

thus $\text{cov}[\mathbf{x}] = \text{diag}(\mu_i(1-\mu_i))$

c) $E x_i = \sum_{k=1}^{K} \pi_k \mu_{ki}$; thus $E\vec{x} = \sum_{k=1}^{K} \pi_k \mu_k$

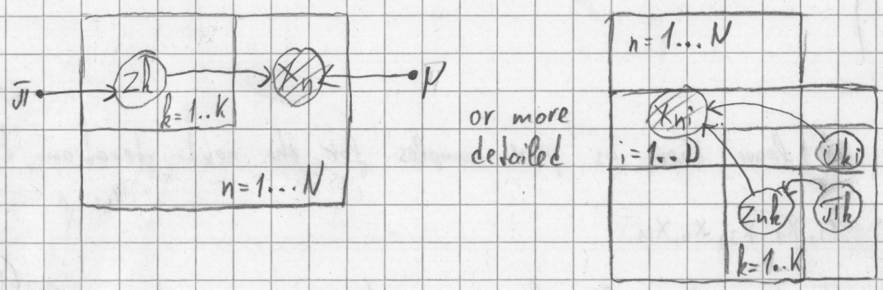
d) $L_I = \prod_{x} p(x|\mu,\pi)$, $\log L_I = \sum_{x}\left[\log \sum_{k=1}^{K}\left[\pi_k \prod_{i=1}^{D} \mu_{ki}^{x_i}(1-\mu_{ki}^{1-x_i})\right]\right]$

e) Because we are dealing with an incomplete log-likelihood, and we have a sum inside of the logarithm, there is no closed form solution and it is challenging to perform the optimization.

f) $\log L_c = \sum_{n=1}^{N}\sum_{k=1}^{K} \log p(x_n, z_{nk}|\mu,\pi)^{z_{nk}} = \sum_{n=1}^{N}\sum_{k=1}^{K}\left[z_{nk}(\log p(z_{nk}|\pi) + \log p(x_n|z_{nk},\mu))\right] =$

$= \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\left(\log \pi_k + \sum_{i=1}^{D}\left[x_{ni}\log \mu_{ki} + (1-x_{ni})\log(1-\mu_{ki})\right]\right)$

g)



h) To find $B(q,\theta)$ (I shall use $\mathcal{L}(q,\theta)$ instead as Bishop does) we take expectation of the complete log-likelihood, which is going to be our lower bound. $\theta = \{\mu, \pi\}$

$$\mathcal{L}(q,\theta) = E_Z[\log L_c] = \sum_{n=1}^{N}\sum_{k=1}^{K} q_n(k)\log\frac{p(x_n, z_{nk}|\mu,\pi)}{q(k)} =$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} q_n(k)\left[\log \pi_k + \sum_{i=1}^{D} x_{ni}\log\mu_{ki} + (1-x_{ni})\log(1-\mu_{ki}) - \log q_n(k)\right] =$$

i) $\tilde{\mathcal{L}}(q,\tilde{\theta}) = \mathcal{L}(q,\theta) + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right) + \sum_{n=1}^{N}\beta_n\left(\sum_{k=1}^{K} q_n(k) - 1\right)$

$\tilde{\theta} = \{\mu, \pi, \lambda, \beta\}$

j) $\frac{d}{d\, q(z_{nk})}\tilde{\mathcal{L}}(q,\tilde{\theta}) = \log\pi_k + \left[\sum_{i=1}^{D} x_{ni}\log\mu_{ki} + (1-x_{ni})\log(1-\mu_{ki})\right] - \log q_n(k) - 1 + \beta_n = 0$

$$q_n(k) = \exp(\beta_n - 1)\cdot\pi_k\cdot\prod_{i=1}^{D}\mu_{ki}^{x_{ni}}(1-\mu_{ki})^{1-x_{ni}}$$

$q_n(k)$ is a normalized posterior probability of assigning datapoint x_n to class k which by investigation is proportional to: $p(z_{nk}|\pi)\,p(x_n|\mu_k)$

k) $\frac{d}{d\pi_k}\tilde{\mathcal{L}}(q,\theta) = \sum_{n=1}^{N}\frac{q_n(k)}{\pi_k} + \lambda = 0 \Rightarrow \pi_k = -\frac{1}{\lambda}\sum_{n=1}^{N} q_n(k)$

$-\pi_k\lambda = \sum_{n=1}^{N} q_n(k) \Rightarrow -\lambda = \sum_{n=1}^{N}\sum_{k=1}^{K} q_n(k) = N$, thus we finally obtain:

$\pi_k = \frac{N_k}{N}$, where $N_k = \sum_{n=1}^{N} q_n(k)$

Problem 3

Because we are sampling from $p(\mu, \tau | x)$, we need $p(\mu | x, \tau)$ and $p(\tau | x, \mu)$

$p(\mu | x, \tau) \propto p(x | \mu, \tau) p(\mu)$ since $x \sim$ Gauss and $\mu \sim$ Gauss, the product will also be Gaussian.

$$p(\mu | x, \tau) \propto C_1 C_2 \exp\left(-\frac{1}{2}\left((x-\mu)^2 \tau + (\mu - \mu_0)^2 s_0^{-1}\right)\right) =$$

$$= C_1 C_2 \exp\left(-\frac{1}{2}\left(x^2\tau - 2x\mu\tau + \mu^2\tau + \mu^2 s_0^{-1} - 2\mu\mu_0 s_0^{-1} + \mu_0^2 s_0^{-1}\right)\right) =$$

$$= C_1 C_2 \exp\left(-\frac{1}{2}\left(\mu^2 \underbrace{(\tau + s_0^{-1})}_{s_N^{-1}} - 2\mu \underbrace{(x\tau + \mu_0 s_0^{-1})}_{s_N^{-1}\mu_N} + \text{const}\right)\right) =$$

And we can instantly see that terms have quadratic structure and thus,

$$p(\mu | x, \tau) = \mathcal{N}\left(\mu \,\middle|\, \frac{x\tau + \mu_0 s_0^{-1}}{\tau + s_0^{-1}}, \frac{1}{\tau + s_0^{-1}}\right) \quad \text{as} \quad \mu_N = (\tau + s_0^{-1})^{-1}(x\tau + \mu_0 s_0^{-1})$$

$$p(\tau | x, \mu) = \frac{p(x | \mu, \tau) p(\tau)}{p(x)} \propto p(x | \mu, \tau) p(\tau) =$$

$$= \frac{1}{\Gamma(a)} b^a \tau^{a-1} \exp(-b\tau) \tau^{1/2} (2\pi)^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)^2 \tau\right) \propto \tau^{a + \frac{1}{2} - 1} \exp\left(-\left(b + \frac{(x-\mu)^2}{2}\right)\tau\right) =$$

Which by inspection we can see is Gamma distribution.

$$\text{Gamma}\left(\tau \,\middle|\, a + \frac{1}{2}, b + \frac{(x-\mu)^2}{2}\right)$$