4. a)

$$P\left[|\mu-\bar{x}|>\varepsilon\right] < \sqrt{1+N}\, e^{\frac{-N^2\varepsilon^2}{2(1+N)}}$$

$$\delta = \sqrt{1+N}\, e^{-N^2\varepsilon^2/2(1+N)}$$

$$\frac{\delta}{\sqrt{1+N}} = e^{-N^2\varepsilon^2/2(1+N)}$$

$$-\ln\left[\frac{\delta}{\sqrt{1+N}}\right] = N^2\varepsilon^2/2(1+N)$$

$$\frac{2(1+N)}{N^2}\ln\left[\frac{\sqrt{1+N}}{\delta}\right] = \varepsilon^2$$

$$\varepsilon = \sqrt{\frac{2(1+N)}{N^2}\ln\left[\frac{\sqrt{1+N}}{\delta}\right]}$$

Now, $$P\left[|\mu-\bar{x}|>\varepsilon\right] < \delta$$

$$\Longleftrightarrow \quad P\left[|\mu-\bar{x}|\leq\varepsilon\right] > 1-\delta$$

$$P\left[|\mu-\bar{x}| \leq \sqrt{\frac{2\cdot(1+N)}{N^2}\ln\left[\frac{\sqrt{1+N}}{\delta}\right]}\right] > 1-\delta$$

where $$-\sqrt{\frac{2(1+N)}{N^2}\ln\left[\frac{\sqrt{1+N}}{\delta}\right]} \leq \mu-\bar{x} \leq \sqrt{\frac{2(1+N)}{N^2}\ln\left[\frac{\sqrt{1+N}}{\delta}\right]}$$

$$\Longrightarrow \quad P\left[\mu \in \left[\bar{x}-\sqrt{\frac{2(1+N)}{N^2}\ln\left[\frac{\sqrt{1+N}}{\delta}\right]},\ \bar{x}+\sqrt{\frac{2(1+N)}{N^2}\ln\left[\frac{\sqrt{1+N}}{\delta}\right]}\right]\right) > 1-\delta$$

b) Derive for Bernoulli first:

→ K actions; when played, prob $\theta_k$ there will be a reward of 1 ($p(r_t=0) = 1-\theta_k$)

$\theta = (\theta_1, \dots, \theta_k)$ are fixed over time.

If we take a beta prior, $\alpha = (\alpha_1, \dots, \alpha_k)$ $\beta = (\beta_1, \dots, \beta_k)$

then $p(\theta_k) = \text{Beta}(\alpha_k, \beta_k)$

The posterior is also Beta. $\alpha_k, \beta_k$ are a count of successes and failures

⇒ Update rule, suppose action $x_t = k$

$$\alpha_k = r_t$$
$$\beta_k = 1 - r_t$$

$r_t \in \{0, 1\}$ reward at time $t$.

Alg:
For $t = 1, 2, \dots$:
  for $k = 1 \dots K$:
    $\theta_k \sim \text{Beta}(\alpha_k, \beta_k)$
  $x_t = \text{argmax}_k \theta_k$
  $r_t \leftarrow \text{apply}(x_t)$
  $(\alpha_{x_t}) = \alpha_{x_t} + r_t$
  $\beta_{x_t} = \beta_{x_t} + (1 - r_t)$

Poisson $\quad x_t = k^{th}$ action

$$p(r_t | x_t, k) = \text{poisson}(\lambda_k)$$

$$p(\lambda_k) = \gamma(\alpha, \beta), \text{ since } \gamma \text{ is poisson conjugate prior.}$$

$$\text{Posterior} \sim \gamma(\alpha + \sum x_i, n + \beta)$$

Algorithm:

for $t = 1, 2, \ldots$

  for $k = 1, \ldots K$

    $\lambda_k \sim \gamma(\alpha_k, \beta_k)$

  $i = \text{argmax}_k \lambda_k$

  $r_t \leftarrow \text{apply}(x_t = i^{th} \text{ action})$

  $\alpha_i = r_t + \alpha_i$

  $\beta_i = 1 + \beta_i$.

Since, for question 4c, we know how the rewards are generated, we can further specify

4cAlg:

for $t = 1, \ldots, 10{,}00$

  for $k = 1, \ldots, K$:

    $\lambda_k \sim \gamma(\alpha_k, \beta_k)$

  $i = \text{argmax}_k \lambda_k$

  $r_t \sim \text{pois}(\text{rate} = i^{th} \text{ actual rate})$ i.e if $x_t = 3$

  $(\alpha_i, \beta_i) = (r_t + \alpha_i, 1 + \beta_i)$ $\quad x_t^* = [1,2,3,4,5][3]$

$= 3$