1. Let $f_{x_i}(X) = \frac{1}{A}, 0 \leq X \leq A$; 0 otherwise ($A$ is unknown). Let $y = max[x_1 \ldots x_n]$.
   $F_y(Y) = Pr[y \leq Y] = Pr[max[x_1 \ldots x_n] \leq Y] = \prod_{i=1}^n Pr[x_i \leq Y] = (\frac{Y}{A})^n, Y \leq A$.
   $f_y(Y) = \frac{d}{dY} F_y(Y) = nY^{n-1}/A^n, 0 \leq Y \leq A$; 0 otherwise (previous result scaled).

1a. $E[\hat{A}] = E[y] = \int_0^A (YnY^{n-1}/A^n)dY = \frac{n}{n+1}A \neq A \rightarrow \hat{A}$ is a *biased* estimator.
   $\underset{n\to\infty}{\text{LIM}} E[\hat{A}] = \underset{n\to\infty}{\text{LIM}} \frac{n}{n+1}A = A \rightarrow \hat{A}$ is an *asymptotically unbiased* estimator.

1b. $Pr[|\hat{A} - A| > \epsilon] = Pr[\hat{A} < A - \epsilon] = \int_0^{A-\epsilon} (nY^{n-1}/A^n)dY = (1 - \frac{\epsilon}{A})^n$.
   $\underset{n\to\infty}{\text{LIM}} Pr[|\hat{A} - A| > \epsilon] = \underset{n\to\infty}{\text{LIM}} (1 - \frac{\epsilon}{A})^n = 0 \rightarrow \hat{A}$ is a (weakly) *consistent* estimator.

2. $f_{\underline{x}|\lambda}(\underline{X}|\lambda) = \prod_{i=1}^5 \lambda e^{-\lambda X_i} = \lambda^5 e^{-\lambda \sum_{i=1}^5 X_i}$ for $X_i \geq 0$.
   $0 = \frac{\partial}{\partial\lambda} \log f_{\underline{x}|\lambda}(\underline{X}|\lambda) = \frac{\partial}{\partial\lambda}[5 \log \lambda - \lambda \sum_{i=1}^5 X_i] = \frac{5}{\lambda} - \sum_{i=1}^5 X_i \rightarrow \hat{\lambda}_{MLE} = 5/\sum_{i=1}^5 X_i$.
   Note that $\sum_{i=1}^5 X_i$ is a *sufficient statistic* for $\hat{\lambda}_{MLE}$: Don't need individual X_i.

3. **Given:** $f_{r|\lambda}(R|\lambda) = \lambda e^{-\lambda R}, R \geq 0.$ $f_\lambda(\lambda) = \frac{1}{T} e^{-\lambda/T}, \lambda \geq 0.$

3a. **MLE:** $0 = \frac{\partial}{\partial\lambda} \log f_{r|\lambda}(R|\lambda) = \frac{\partial}{\partial\lambda}[\log \lambda - \lambda R] = \frac{1}{\lambda} - R \rightarrow \hat{\lambda}_{MLE}(R) = \frac{1}{R}$. Note units.

3b. **MAP:** $0 = \frac{\partial}{\partial\lambda}[\log f_{r|\lambda}(R|\lambda) + \log f_\lambda(\lambda)] = \frac{\partial}{\partial\lambda}[\log \lambda - \lambda R - \log T - \lambda/T] = \frac{1}{\lambda} - R - \frac{1}{T} \rightarrow$
   $\hat{\lambda}_{MAP}(R) = 1/(R + 1/T) = T/(RT + 1)$=1/(a priori+a posteriori).

   $(T \rightarrow 0) \rightarrow \lambda = 0 \rightarrow \hat{\lambda}_{MAP}(R) \rightarrow 0$ (a priori information about λ predominates).
   $(T \rightarrow \infty) \rightarrow f_\lambda(\lambda)$ uniform$\rightarrow \hat{\lambda}_{MAP}(R) \rightarrow \hat{\lambda}_{MLE}(R)$ (a posteriori info predominates).

3c. **MSE:** $\hat{\lambda}_{LSE}(R) = \frac{\int \lambda f_{r|\lambda}(R|\lambda) f_\lambda(\lambda) d\lambda}{\int f_{r|\lambda}(R|\lambda) f_\lambda(\lambda) d\lambda} = \frac{\int \frac{\lambda^2}{T} e^{-\lambda(R+1/T)} d\lambda}{\int \frac{\lambda}{T} e^{-\lambda(R+1/T)} d\lambda} = 2/(R + 1/T) = 2\hat{\lambda}_{MAP}$.

   NOTE: $f_{\lambda|R}(\lambda|R) = (R + 1/T)^2 \lambda e^{-\lambda(R+1/T)}, \lambda > 0 = 2^{nd}$-order Erlang pdf.
   Recognizing this, read off $\hat{\lambda}_{LSE}(R) = E[\lambda|r = R] = 2/(R + 1/T)$ without integrals.

4. $p_{x_1 \ldots x_{100}|p}(X_1 \ldots X_{100}|P) = P^K (1 - P)^{100-K}$ where $K = \sum_{i=1}^{100} X_i$=#heads.

4a. $\frac{\partial}{\partial P}[K \log P + (100 - K) \log(1 - P)] = \frac{K}{P} - \frac{100-K}{1-P} \rightarrow \hat{p}_{MLE} = \frac{K}{100}$.

4b. $\hat{p}_{LLSE}(K) = E[p] + \frac{\lambda_{kp}}{\sigma_k^2}(K - E[k])$. Need to plug into formula:

   $E[p] = \frac{1}{2}$. $E[k] = E_p[E[k|p]] = E[100p] = 100(\frac{1}{2}) = 50$.
   $E[kp] = E_p E[kp|p] = E[100p^2] = 100 \int_0^1 P^2 dP = 33.33$.
   $\lambda_{kp} = E[kp] - E[k]E[p] = 33.33 - 50(\frac{1}{2}) = 8.33$.
   $\sigma_k^2 = E_p[\sigma_{k|p}^2] + Var_p[E[k|p]] = E[100p(1 - p)] + \sigma_{100p}^2 = \frac{100}{2} - \frac{100}{3} + 10^4 \frac{(1-0)^2}{12} = 850$.

   $\hat{p}_{LLSE}(K) = \frac{1}{2} + \frac{8.33}{850}(K - 50) = 0.009804(K + 1) = \frac{K+1}{102}$ (look familiar?).
   NOTE: Can also compute $\hat{p}_{LSE}(K) = \frac{K+1}{102}$ and note this is a *linear* estimator!