Example Conceptual Problems

There are a huge variety of possible conceptual problems. Here are a few examples.

Problem 1. A 90% confidence interval for the average number of children per household based on a simple random sample is found to be (0.7, 2.1). Because the average number of children per household, μ, is some fixed number in the population (at least, at a particular moment in time when we conduct the study), it doesn't make any sense to claim that $P(0.7 \leq \mu \leq 2.1) = 0.90$. What do we mean, then, by saying that this is a "90% confidence interval"? Can we ever make probability statements about confidence intervals?

For 90% of samples, an interval calculated in this way will contain the population mean μ.

Before taking the sample, the interval endpoints are random variables; denote them by A and B. Then the interval [A, B] is a random interval, and we can write

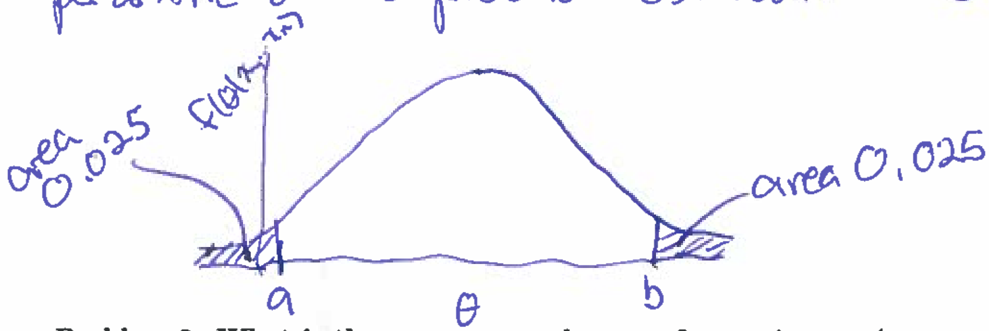
$$P(A \leq \mu \leq B) = 0.90$$

We cannot make probability statements about the realized values of random variables, so once we have taken a sample and observed $a = 0.7$, $b = 2.1$, it does not make sense to write

$$P(0.7 \leq \mu \leq 2.1) = 0.90.$$

Problem 2. What is a posterior distribution in a Bayesian analysis? If I know the posterior distribution for a model parameter, how can a 95% posterior credible interval be formed? (You should answer the first question with a written sentence. For the second question, you could write a sentence and/or draw a picture to illustrate.)

A posterior is a probability distribution that represents our state of knowledge about a parameter after having observed the sample data. A 95% posterior credible interval is an interval $[a, b]$ such that the posterior assigns probability 0.95 to that ~~inter~~ interval : $P(\Theta \in [a,b] \mid x_1, \ldots, x_n) = 0.95$.
One way to achieve this is by setting a to be the 2.5th percentile of the posterior distribution and b the 97.5th percentile.



Problem 3. What is the mean squared error of an estimator (you can answer with either a formula or a written sentence explaining the intuition)? Why is an estimator with low mean squared error preferred to an estimator with high mean squared error?

$MSE(\hat{\theta}) = E[\{\hat{\theta} - \theta\}^2]$ is the average squared difference between the estimate and the parameter being estimated. We would like our estimates to be close to the parameter being estimated on average, which corresponds to a low MSE.

Example Worked Problems

The midterm will have problems roughly similar in content to the examples below.

Problem 1

The EPA conducts occasional reviews of its standards for airborne asbestos. During a review, the EPA examines data from several studies (denote the number of studies by s). Different studies keep track of different groups of people; different groups have different exposures to asbestos. Let n_i be the number of people in the i'th study, let x_i be the asbestos exposure for people in that study, and let y_i be the number of people who developed lung cancer in that study. The EPA's model is $Y_i \sim \text{Poisson}(\lambda_i)$, where $\lambda_i = n_i x_i \lambda$ and where λ is the typical rate at which asbestos causes cancer. The n_i's and x_i's are known constants; the Y_i's are random variables. Because the different studies involve different sets of people in different locations, they model the Y_i's from different studies as being independent (but not identically distributed since the λ_i's are different!). The EPA wants to estimate λ.

In answering the questions below, you may use the following facts about the Poisson and Gamma distributions:

Suppose $X \sim \text{Poisson}(\lambda)$

p.f.	$f(x\|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$
Mean	λ
Variance	λ

$$\to f(y_i|\lambda_i) = e^{-\lambda_i}\frac{\lambda_i^{y_i}}{y_i!}$$

$$= e^{-\lambda n_i x_i}\frac{(\lambda n_i x_i)^{y_i}}{y_i!}$$

Suppose $X \sim \text{Gamma}(\alpha, \beta)$

p.f.	$f(x\|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$
Mean	$\frac{\alpha}{\beta}$
Variance	$\alpha\beta^2$

(a) Find the pdf of the joint distribution of $Y_1, \ldots, Y_s | x_1, n_1, \ldots, x_s, n_s, \lambda$.

$$f_{Y_1,\ldots,Y_s|\lambda}(y_1,\ldots,y_s|\lambda) = \prod_{i=1}^{n} f_{Y_i|\lambda}(y_i|\lambda)$$

$$= \prod_{i=1}^{s} e^{-\lambda n_i x_i}\frac{(\lambda n_i x_i)^{y_i}}{y_i!}$$

$$= \prod_{i=1}^{s} e^{-\lambda n_i x_i}\lambda^{y_i}\cdot\left\{\frac{(n_i x_i)^{y_i}}{y_i!}\right\}$$

$$= e^{-\lambda\sum_{i=1}^{s} n_i x_i}\cdot\lambda^{\sum_{i=1}^{s} y_i}\cdot\prod_{i=1}^{s}\left\{\frac{(n_i x_i)^{y_i}}{y_i!}\right\}$$

(b) Find the maximum likelihood estimator of λ.

Continuing from (a), the log-likelihood is

$$\ell(\lambda | y_1, \ldots, y_s) = -\lambda \cdot \sum_{i=1}^{s} n_i x_i + \sum_{i=1}^{s} y_i \cdot \log(\lambda) + \log\left[\prod_{i=1}^{s}\left\{\frac{(n_i x_i)^{y_i}}{y_i!}\right\}\right]$$

The first and second derivatives are

$$\frac{d}{d\lambda}\ell(\lambda | y_1, \ldots, y_s) = -\sum_{i=1}^{s} n_i x_i + \frac{1}{\lambda} \cdot \sum_{i=1}^{s} y_i$$

$$\frac{d^2}{d\lambda^2}\ell(\lambda | y_1, \ldots, y_s) = \frac{-1}{\lambda^2}\sum_{i=1}^{s} y_i$$

Setting the first derivative to 0 and solving for λ we obtain

$$0 = -\sum_{i=1}^{s} n_i x_i + \frac{1}{\lambda}\sum_{i=1}^{s} y_i$$

$$\Rightarrow \lambda = \frac{\sum_{i=1}^{s} y_i}{\sum_{i=1}^{s} n_i x_i}$$

Since the second derivative is negative, this is a global maximum.

The maximum likelihood estimator is

$$\hat{\lambda}^{MLE} = \frac{\sum_{i=1}^{s} Y_i}{\sum_{i=1}^{s} n_i x_i}$$

note use of capital Y_i!

(c) Is the maximum likelihood estimator an unbiased estimator of λ?

$$E[\hat{\lambda}^{MLE}] = E\left[\frac{\sum_{i=1}^{s} Y_i}{\sum_{i=1}^{s} n_i x_i}\right] = \frac{1}{\sum_{i=1}^{s} n_i x_i} \cdot \sum_{i=1}^{s} E[Y_i] = \frac{1}{\sum_{i=1}^{s} n_i x_i} \sum_{i=1}^{s} \lambda_i$$

$$= \frac{1}{\sum_{i=1}^{s} n_i x_i} \sum_{i=1}^{s} \lambda n_i x_i = \lambda \cdot \frac{1}{\sum_{i=1}^{s} n_i x_i} \cdot \sum_{i=1}^{s} n_i x_i = \lambda.$$

Yes, $\hat{\lambda}^{MLE}$ is an unbiased estimator of λ.

(d) Find the variance of the maximum likelihood estimator.

$$Var(\hat{\lambda}^{MLE}) = Var\left(\frac{\sum_{i=1}^{s} Y_i}{\sum_{i=1}^{s} n_i x_i}\right) = \left(\frac{1}{\sum_{i=1}^{s} n_i x_i}\right)^2 \cdot \sum_{i=1}^{s} Var(Y_i)$$

$$= \left(\frac{1}{\sum_{i=1}^{s} n_i x_i}\right)^2 \cdot \sum_{i=1}^{s} \lambda n_i x_i = \frac{1}{\sum_{i=1}^{s} n_i x_i} \cdot \lambda$$

(e) Find the mean squared error of the maximum likelihood estimator.

$$MSE(\hat{\lambda}^{MLE}) = \{Bias(\hat{\lambda}^{MLE})\}^2 + Var(\hat{\lambda}^{MLE})$$

$$= 0^2 + \frac{1}{\sum_{i=1}^{s} n_i x_i} \lambda$$

$$= \frac{1}{\sum_{i=1}^{s} n_i x_i} \cdot \lambda$$

(f) Suppose the EPA uses this model to estimate λ by combining data from $s = 3$ studies with data recorded in the table below. Find an expression for the maximum likelihood estimate of λ. Your answer should involve only numbers, no symbols; but you do not need to simplify your expression.

Study Number (i)	Sample Size (n_i)	Exposure Level (x_i)	Cancer Case Count (y_i)
1	10	0.3	1
2	25	0.2	3
3	100	0.5	15

$$\hat{\lambda}^{MLE} = \frac{1+3+15}{10 \cdot 0.3 + 25 \cdot 0.2 + 100 \cdot 0.5}$$

(g) Suppose the analysts adopt a prior of $\Lambda \sim$ Gamma(α, β), where α and β are known constants they choose to reflect their prior knowledge about λ. Find the posterior distribution for Λ. You should arrive at a specific form for the posterior distribution, with parameters involving α, β, $x_1, \ldots, x_s$, and $n_1, \ldots, n_s$. and $y_1, \ldots, y_s$

$$f_{\Lambda|Y_1,\ldots,Y_s}(\lambda|y_1,\ldots,y_s) \propto f_\Lambda(\lambda)\cdot f_{Y_1,\ldots,Y_s|\Lambda}(y_1,\ldots,y_s|\lambda)$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}\cdot e^{-\lambda\sum_{i=1}^{s} n_i x_i}\cdot\lambda^{\sum_{i=1}^{s} y_i}\cdot\prod_{i=1}^{s}\left\{\frac{(n_i x_i)^{y_i}}{y_i!}\right\}$$

$$\propto \lambda^{\alpha+\sum_{i=1}^{s} y_i - 1} e^{-(\beta+\sum_{i=1}^{s} n_i x_i)\lambda}$$

This is proportional to the pdf of a Gamma distribution.

The posterior is

$$\Lambda|Y_1,\ldots,Y_s \sim \text{Gamma}\left(\alpha+\sum_{i=1}^{s} y_i,\ \beta+\sum_{i=1}^{s} n_i x_i\right)$$

(h) Again, the EPA uses this model to estimate λ by combining data from three studies with data recorded in the table below. They use a prior of $\Lambda \sim$ Gamma(1, 3). Find expressions for the parameters of the posterior distribution for Λ. Your answer should involve only numbers, no symbols; but you do not need to simplify your expression.

Study Number (i)	Sample Size (n_i)	Exposure Level (x_i)	Cancer Case Count (y_i)
1	10	0.3	1
2	25	0.2	3
3	100	0.5	15

The first parameter of the posterior is

$\alpha^{post} = 1+1+3+15$

The second parameter of the posterior is

$\beta^{post} = 3 + 10\cdot 0.3 + 25\cdot 0.2 + 100\cdot 0.5$

Problem 2. From independent surveys of two populations, 90% confidence intervals for the population means μ_1 and μ_2 will be constructed. Denote the first interval, which is an estimate of μ_1, by $[L_1, U_1]$ and the second interval, which is an estimate of μ_2, by $[L_2, U_2]$. We have not taken our sample yet, so L_1, U_1, L_2, and U_2 are random variables. What is the probability that both of these confidence intervals will contain their respective population means?

$$P(L_1 \le \mu_1 \le U_1 \text{ and } L_2 \le \mu_2 \le U_2)$$

$$= P(L_1 \le \mu_1 \le U_1) \cdot P(L_2 \le \mu_2 \le U_2)$$

$$= 0.9 \cdot 0.9$$

$$= 0.81$$

Problem 3. Two suveys were independently conducted to estimate a population mean μ. Denote the estimators from the independent surveys and their variances by $\hat{\mu}_1$, with variance $\sigma_1^2 > 0$ and $\hat{\mu}_2$, with variance $\sigma_2^2 > 0$. Assume that both $\hat{\mu}_1$ and $\hat{\mu}_2$ are unbiased. For some constants α and β, the two estimators can be combined to give a new estimator $Y = \alpha\hat{\mu}_1 + \beta\hat{\mu}_2$.

(a) Find a condition on α and β so that the combined estimator Y is unbiased.

We need $E[Y] = \mu$:

$$E[Y] = E[\alpha\hat{\mu}_1 + \beta\hat{\mu}_2] = \alpha \cdot E[\hat{\mu}_1] + \beta \cdot E[\hat{\mu}_2] = \alpha\mu + \beta\mu$$

$$\mu(\alpha+\beta) = \mu$$

$$\alpha + \beta = 1$$

(b) What choice of α and β minimizes the variance of Y, subject to the condition of unbiasedness?

$$Var(\alpha\hat{\mu}_1 + \beta\hat{\mu}_2) = \alpha^2\sigma_1^2 + \beta^2\sigma_2^2$$

Use the condition that $\alpha+\beta=1$, or $\alpha = 1-\beta$:

$$Var(Y) = (1-\beta)^2\sigma_1^2 + \beta^2\sigma_2^2 = (1-2\beta+\beta^2)\sigma_1^2 + \beta^2\sigma_2^2$$

Find β to minimize: take the first derivative and set to 0:

$$\frac{d}{d\beta} Var(Y) = -2\sigma_1^2 + 2\beta\sigma_1^2 + 2\beta\sigma_2^2 = 0$$

$$\Rightarrow \beta(\sigma_1^2 + \sigma_2^2) = \sigma_1^2$$

$$\Rightarrow \beta = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}, \text{ so } \alpha = 1-\beta = 1 - \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

The second derivative is $\frac{d^2}{d\beta^2} Var(Y) = 2\sigma_1^2 + 2\sigma_2^2 > 0$,

so the critical point above is a global minimum of $Var(Y)$