1. Take \( \{X_n\} \) such that \( P(X_n = n) = 1/n \) and \( P(X_n = 0) = 1 - 1/n \). Then \( P(X_n > 0) = 1/n \to 0 \) so \( X_n \xrightarrow{p} 0 \); however, \( E(X_n) = 0 \cdot (1 - 1/n) + n \cdot 1/n = 1 \not\to 0 \).

2. The \( X_i \)'s are independent (but not iid) so the expected value and variance of \( \sum_{i=1}^n X_i \) is the sum of the individual expected values and variances; in fact, because the distribution is Poisson, the means and variances are the same:

\[
\sum_{i=1}^n E(X_i) = \sum_{i=1}^n V(X_i) = \sum_{i=1}^n \theta_i.
\]

Take any \( \varepsilon > 0 \). By Chebyshev’s inequality we get

\[
P\left( \left| \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n \theta_i} - 1 \right| > \varepsilon \right) = P\left( \left| \sum_{i=1}^n X_i - \sum_{i=1}^n \theta_i \right| > \varepsilon \sum_{i=1}^n \theta_i \right)
\leq \frac{\sum_{i=1}^n \theta_i}{\varepsilon^2(\sum_{i=1}^n \theta_i)^2} = \frac{1}{\varepsilon^2 \sum_{i=1}^n \theta_i}.
\]

Since \( \sum_{i=1}^n \theta_i \to \infty \), the right-hand side goes to 0; therefore, \( \sum_{i=1}^n X_i / \sum_{i=1}^n \theta_i \xrightarrow{p} 1 \).

3. By the central limit theorem, \( \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2) \). Consider the function \( g(x) = x^2 \). Since \( g'(\mu) \neq 0 \), the delta theorem says that the asymptotic distribution of \( g(\bar{X}_n) = \bar{X}_n^2 \) is \( \sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{d} N(0, 4\mu^2\sigma^2) \).

4. For the geometric distribution with success probability \( \theta \) and support \( \{1, 2, 3, \ldots\} \), recall that \( E_\theta(X) = 1/\theta \) and \( V_\theta(X) = (1 - \theta)/\theta^2 \); see, e.g., wikipedia.

(a) The likelihood function is

\[
L(\theta) = \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} = \theta^n (1 - \theta)^{n(\bar{x} - 1)}.
\]

Taking log and then derivative gives

\[
0 = \frac{\partial}{\partial \theta} \log L(\theta) = \frac{\partial}{\partial \theta} \left[ n \log \theta + n(\bar{x} - 1) \log(1 - \theta) \right] = \frac{n}{\theta} - \frac{n(\bar{x} - 1)}{1 - \theta}.
\]

Using basic algebra it can be deduced that \( \hat{\theta}_n = 1/\bar{X}_n \). The law of large numbers says that \( \bar{X}_n \xrightarrow{p} E_\theta(X_1) = 1/\theta \), so \( \hat{\theta}_n \xrightarrow{p} \theta \).

(b) The central limit theorem gives \( \sqrt{n}(\bar{X}_n - 1/\theta) \xrightarrow{d} N(0, (1 - \theta)/\theta^2) \). Use the delta theorem with \( g(x) = 1/x \) and the fact that \( \hat{\theta}_n = g(\bar{X}_n) \) to conclude that

\[
\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, [g'(1/\theta)]^2 \cdot (1 - \theta)/\theta^2\right) = N(0, \theta^2(1 - \theta)).
\]
(c) Simple calculus gives
\[ \frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = -\frac{1}{\theta^2} - \frac{x-1}{(1-\theta)^2}. \]
Taking expected value gives the Fisher information:
\[ I(\theta) = \mathbb{E}_\theta \left[ -\frac{\partial^2}{\partial \theta^2} \log f_\theta(X) \right] = \frac{1}{\theta^2} + \frac{\mathbb{E}_\theta(X) - 1}{(1-\theta)^2} = \frac{1}{\theta^2} + \frac{1/\theta - 1}{(1-\theta)^2} = \frac{1}{\theta^2(1-\theta)}. \]

(d) The asymptotic distribution of the MLE is \( \sqrt{n}(\hat{\theta}_n - \theta) \overset{d}{\rightarrow} \mathcal{N}(0, \theta^2(1-\theta)) \). This is the same as in part (b). In fact, it is not possible for a sequence of random variables to have more than one asymptotic distribution.

(e) By definition \( P_\theta(X > 7) \) is the sum of the PMF \( f_\theta(x) \) for all \( x > 7 \). That is,
\[ P_\theta(X > 7) = \sum_{x=7+1}^{\infty} \theta(1-\theta)^{x-1} = (1-\theta)^7 \sum_{y=1}^{\infty} \theta(1-\theta)^{y-1} = (1-\theta)^7. \]
(The key to this calculation is that there’s a common factor of \( (1-\theta)^7 \) which can be pulled out front of the summation. Then by making a change-of-variable \( (y = x - 7) \) the summation looks exactly like the summation of the PMF over all possible \( x \) values, which we know to be 1.) Now let \( g(\theta) = P_\theta(X > 7) = (1-\theta)^7 \). According to Theorem 6.1.2 in the text, the MLE of \( g(\theta) \) is \( g(\hat{\theta}_n) = (1 - 1/\bar{X}_n)^7 \). Use the delta theorem to get the asymptotic distribution:
\[ \sqrt{n}[g(\hat{\theta}_n) - g(\theta)] \overset{d}{\rightarrow} \mathcal{N}(0, 49\theta^2(1-\theta)^{14}). \]

5. (a) This is the same as Problem 6.1.2(b) in HMC7. The MLE is \( \hat{\theta} = X_{(1)} \), the sample minimum. This calculation required the use of indicator functions, and so the usual differentiation of the log-likelihood wasn’t possible. See my solutions to that homework problem.

(b) First focus on \( \hat{\theta} = X_{(1)} \). Using independence, it is easy to check that
\[ P_\theta(X_{(1)} > x) = P_\theta(X_1 > x, \ldots, X_n > x) = P_\theta(X_1 > x)^n = e^{-n(x-\theta)}. \]
Then the CDF of \( n(\hat{\theta} - \theta) \) is
\[ P_\theta(n(\hat{\theta} - \theta) \leq z) = P_\theta(\hat{\theta} \leq \theta + z/n) = 1 - e^{-n(\theta + z/n - \theta)} = 1 - e^{-z}. \]
This is the \( \exp(1) \) CDF, so we conclude \( n(\hat{\theta} - \theta) \sim \exp(1) \).

(c) The expected value of \( \exp(1) \) is 1; therefore, from part (b), \( 1 = E_\theta[n(\hat{\theta} - \theta)] = nE(\hat{\theta}) - n\theta \), which implies \( E_\theta(\hat{\theta}) = \theta + 1/n \). So \( \hat{\theta} \) is not unbiased.

(d) To calculate the MSE of \( \hat{\theta} \), use the bias–variance decomposition. That is,\[ \text{MSE}_\theta(\hat{\theta}) = \text{Var}_\theta(\hat{\theta}) + b_\theta(\hat{\theta})^2. \]
The bias term, in this case, is \( b_\theta(\hat{\theta})^2 = 1/n^2 \) from the calculation in part (c). For the variance of \( \hat{\theta} \), we can use part (b) again. That is, since the variance of \( \exp(1) \) is 1, we get \( 1 = \text{Var}[n(\hat{\theta} - \theta)] = n^2\text{V}(\hat{\theta}) \), i.e., \( \text{V}(\hat{\theta}) = 1/n^2 \). Putting everything together, we find that the MSE of \( \hat{\theta} \) is \( 2/n^2 \).
6. (a) As shown in class, the MLE of $\theta$ is $\hat{\theta}_n = \bar{X}_n$. Since $\theta(1 - \theta)$ is a function of $\theta$, its MLE is just $\bar{X}_n(1 - \bar{X}_n)$.

(b) To check for unbiasedness,

$$E_{\theta}[\bar{X}_n(1 - \bar{X}_n)] = E_{\theta}(\bar{X}_n) - E_{\theta}(\bar{X}_n^2) = \theta - \chi_{\theta}(\bar{X}_n) - E_{\theta}(\bar{X}_n)^2$$

$$= \theta - \theta(1 - \theta) - \theta^2 = \frac{n - 1}{n} \cdot \theta(1 - \theta).$$

So $\bar{X}_n(1 - \bar{X}_n)$ is not unbiased, although it’s close to unbiased when $n$ is large. But consistency follows from the continuous mapping theorem.

7. (a) $Y_1, \ldots, Y_n \sim \text{Ber}(\eta)$, where $\eta = P_\theta(X > 7) = e^{-7/\theta}$. So $\theta$ and $\eta$ are related: $\eta = g(\theta) = e^{-7/\theta}$ or, equivalently, $\theta = g^{-1}(\eta) = -7/\log \eta$.

(b) From the corrupted sample, the MLE of $\eta$ is $\hat{\eta}_n = \bar{Y}_n$. Since $\theta$ is a function of $\eta$, the MLE of $\theta$ is $\hat{\theta}_n = -7/\log \bar{Y}_n$.

(c) Consistency of $\hat{\theta}_n$ follows from the LLN and continuous mapping theorem.

(d) Let $I_Y(\theta)$ denote the information about $\theta$ in a sample $Y \sim \text{Ber}(e^{-7/\theta})$. Then the Cramer–Rao lower bound for estimating $\theta$ based on $Y_1, \ldots, Y_n \sim \text{Ber}(e^{-7/\theta})$ is $[nI_Y(\theta)]^{-1}$. To calculate $I_Y(\theta)$, first let $f_\eta(y) = \eta^y(1-\eta)^{1-y}$ where $\eta = e^{-7/\theta}$. The key is that differentiation with respect to $\theta$ can be carried out via the chain rule. This is made clear in the following expression for $I_Y(\theta)$:

$$I_Y(\theta) = E_{\theta} \left[ \left( \frac{\partial}{\partial \eta} \log f_{\eta=\exp(-7/\theta)}(Y) \right)^2 \right]$$

$$= \left( \frac{d\eta}{d\theta} \right)^2 E_{\theta} \left[ \left( \frac{\partial}{\partial \eta} \log f_\eta(Y) \right) \bigg|_{\eta=\exp(-7/\theta)} \right]^2.$$

In this Bernoulli case, the expression becomes

$$I_Y(\theta) = \frac{(49/\theta^4)e^{-14/\theta}}{e^{-7/\theta}(1 - e^{-7/\theta})} = \frac{1}{\theta^2} \cdot \frac{49e^{-7/\theta}}{\theta^2(1 - e^{-7/\theta})}.$$

The second term, $h(\theta)$, in the above product is strictly less than 1 as a function of $\theta$; see Figure 1. On the other hand, the Fisher information $I_X(\theta)$ for $\theta$ in a single sample $X$ from the original population is $I_X(\theta) = 1/\theta^2$. Therefore,

$$I_Y(\theta) = \frac{1}{\theta^2} \cdot h(\theta) < \frac{1}{\theta^2} = I_X(\theta).$$

So Fisher information agrees with the interpretation of data corruption as “loss of information.”

8. (a) The MGF for $X \sim \text{Pois}(\theta)$ is $M_X(s) = E_{\theta}(e^{sX}) = e^{\theta(e^s-1)}$. If $T = X_1 + \cdots + X_n$ and the $X_i$’s are independent, then the MGF of $T$ is

$$M_T(s) = E_{\theta}(e^{sT}) = E_{\theta}(e^{s(X_1+\cdots+X_n)}) = \prod_{i=1}^{n} E_{\theta}(e^{sX_i}) = \prod_{i=1}^{n} M_{X_i}(s) = e^{n\theta(e^s-1)}.$$

This is the MGF for $\text{Pois}(n\theta)$, and by the uniqueness of MGFs (Theorem 1.9.1 in HMC7), we conclude that the distribution of $T$ must be $\text{Pois}(n\theta)$. 

3
Figure 1: Plot of the function $h(\theta)$ in Problem #7(d).

(b) For $\hat{\eta}_n = (1 - \frac{1}{n})^T$, the expected value is

$$E_{\theta}(\hat{\eta}_n) = \sum_{t=0}^{\infty} (1 - \frac{1}{n})^t e^{-n\theta} \frac{(n\theta)^t}{t!} = \sum_{t=0}^{\infty} e^{-n\theta} \frac{(n\theta - \theta)^t}{t!} = e^{-\theta} e^{n\theta} = e^{-\theta}.$$

Therefore, $\hat{\eta}_n$ is an unbiased estimator of $\eta = e^{-\theta}$.

(c) First find $E_{\theta}(\hat{\eta}_n^2)$:

$$E_{\theta}(\hat{\eta}_n^2) = \sum_{t=0}^{\infty} (1 - \frac{1}{n})^{2t} e^{-n\theta} \frac{(n\theta)^t}{t!} = \sum_{t=0}^{\infty} e^{-n\theta} \frac{(n\theta(1 - \frac{1}{n})^2)^t}{t!} = e^{-n\theta+n\theta(1-\frac{1}{n})^2}.$$

So then $V_{\theta}(\hat{\eta}_n) = e^{-2\theta+\theta/n} - e^{-2\theta}$. Since the variance is clearly vanishing as $n \to \infty$, Chebyshev’s inequality implies $\hat{\eta}_n$ is consistent.

(d) The Cramer–Rao lower bound for estimating $\eta = e^{-\theta}$ is

$$LB = \frac{(d\eta/d\theta)^2}{nI(\theta)} = \frac{\theta e^{-2\theta}}{n}.$$

Then the efficiency of $\hat{\eta}_n$ is the ratio LB by $V_{\theta}(\hat{\eta}_n)$:

$$\text{eff}_{\theta}(\hat{\eta}_n) = \frac{LB}{V_{\theta}(\hat{\eta}_n)} = \frac{\theta e^{-2\theta/n}}{e^{-2\theta}[e^{\theta/n} - 1]} = \frac{\theta/n}{e^{\theta/n} - 1} = \left(\frac{e^{\theta/n} - 1}{\theta/n}\right)^{-1}.$$

This is less than 1 for all fixed $n$; however, the limit as $n \to \infty$ is one so $\hat{\eta}_n$ is asymptotically efficient.