CMSC 35400 / STAT 37710

Spring 2020
Homework 2

1. Suppose $x_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$, $i = 1, \ldots, n$.

a) Find the method of moments estimate (MOME) $\hat{\theta}_{\text{MOME}}$ of θ based on the first moment.

SOLUTION: The first moment of the Bernoulli distribution is $\mu_1 = \mathbb{E}[X] = \theta$. Therefore, plugging in the sample moment as the estimator for μ_1, the method of moment estimator based on the first moment is

$$\hat{\theta}_{\text{MOME}} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

b) Find the MOME of θ based on the second moment.

SOLUTION: The second theoretical moment about the origin is $\mu_2 = \mathbb{E}_{X\sim p_\theta}[X^2] = \theta$. Therefore, plugging in the second sample moment as the estimator for μ_2, we get

$$\hat{\theta}_{\text{MOME}} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Note that since $x^j = x$ for every j, it follows that $\mu_j = \theta$ for every j. So any of the method of moments equations would lead to the sample mean as the estimator of θ.

c) Find the maximum likelihood estimator (MLE) of θ.

SOLUTION: Recall that

$$p(x|\theta) = \theta^{s(x)}(1-\theta)^{n-s(x)}$$

where $s(x) = \sum_{i=1}^{n} x_i$. Differentiating we find that

$$\frac{\partial}{\partial\theta} p(x|\theta) = \theta^{s(x)-1}(1-\theta)^{n-1-s(x)}[s(x) - n\theta],$$

which is equal to zero if

$$s(x) - n\theta = 0 \quad \Rightarrow \quad \hat{\theta}_{\text{MLE}} = \frac{s(x)}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

d) Find the MLE of the variance of X.

SOLUTION: Similar to problem 1, this is another application of the invariance principle of the MLE. We know that the variance of a Bernoulli random variable is

$$\text{Var}(X) = \theta(1-\theta),$$

therefore we first find the MLE of θ, the parameter of the Bernoulli distribution, which is the sample average. Therefore we find that the MLE of the variance is

$$\widehat{\mathrm{Var}}_{\mathrm{MLE}}(x) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(1 - \frac{1}{n}\sum_{i=1}^{n} x_i\right).$$

2. Suppose that we are studying computer network traffic, and that we are interested in estimating the probability that the traffic rate is less than 13 packets per 10^{-3} second. To estimate this probability, we count the number of packets crossing the network in a 10^{-3} second interval at 20 different times. Assume that these counts independent and identically Poisson distributed (where the density function is given by $p(x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$). The packet counts are:

11, 7, 5, 9, 4, 8, 7, 7, 8, 5, 6, 7, 6, 12, 6, 7, 10, 6, 7, 6

a) Find the MLE for the parameter of the Poisson distribution.

SOLUTION:

$$\hat{\lambda}_{\mathrm{MLE}} = \arg\max_{\lambda} p(x|\lambda) = \arg\max_{\lambda} \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \arg\max_{\lambda} \sum_{i=1}^{n} -\lambda + x_i \log\lambda.$$

Differentiating:

$$\frac{\partial}{\partial\lambda}\left(\sum_{i=1}^{n} -\lambda + x_i \log\lambda\right) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i,$$

and equating to zero yields that

$$\hat{\lambda}_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{20}\sum_{i=1}^{20} x_i = \frac{144}{20} = 7.2.$$

b) Find the MLE for the probability of interest.

SOLUTION: The key is to use the invariance principle of the maximum likelihood estimator. This invariance principle lets us use this estimate to compute the quantity of interest. We would like to estimate

$$p = \mathbb{P}(X < 13|\lambda),$$

and the invariance principle tells us that we can estimate this quantity by simply using $\hat{\lambda}_{\mathrm{MLE}}$:

$$\hat{p}_{\mathrm{MLE}} = \mathbb{P}(X < 13|\hat{\lambda}_{\mathrm{MLE}}) = \sum_{k=0}^{12} \frac{\hat{\lambda}_{\mathrm{MLE}}^{k}\exp(-\hat{\lambda}_{\mathrm{MLE}})}{k!} = 0.9673.$$

c) Suppose that you believe the Gamma distribution

$$p(\lambda) := \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

is a good prior for the Poisson parameter λ, where $\Gamma(\cdot)$ is the Gamma function, and you also know the values of $\alpha = \beta = 5$. Derive the Maximum a Posteriori (MAP) estimate $\hat{\lambda}_{\text{MAP}}$.

SOLUTION:

$$\begin{aligned} p(\lambda \mid x) &\propto p(x \mid \lambda) p(\lambda) \\ &\propto \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\ &\propto \lambda^{\alpha+n\bar{x}-1} e^{-(\beta+n)\lambda} \end{aligned}$$

which is still a Gamma distribution. Therefore, the log posterior is

$$\log p(\lambda \mid x) \propto (\alpha + n\bar{x} - 1)\log\lambda - (\beta + n)\lambda$$

Set the derivative to 0:

$$0 = \frac{\alpha + n\bar{x} - 1}{\lambda} - (\beta + n)$$

We therefore get

$$\lambda = \frac{\alpha + n\bar{x} - 1}{\beta + n} = \frac{5 + 144 - 1}{5 + 20} = 5.92$$

3. Suppose we are monitoring credit card payments for a population of n people, and model the number of days (τ) that will pass before each person defaults on their payments as independent, identically exponentially distributed random variables with parameter $\theta > 0$. That is, for $i = 1, \ldots, n$, $\tau_i \sim \exp(\theta)$, so that $p(\tau_i|\theta) = \theta e^{-\tau_i\theta}$. If a new person applies for a credit card, and we assume that his/her time to default follows the same distribution as the n previous people we've monitored, and is independent of those people, what is the maximum likelihood estimate of the probability of this person defaulting in less than 10 days? (HINT: $\int_a^b te^{-xt}dx = e^{-ta} - e^{-tb}$.)

SOLUTION:

$$\begin{aligned} p(\tau|\theta) &= \prod_{i=1}^{n} \theta e^{-\tau_i\theta} \\ -\log p(\tau|\theta) &= -n\log\theta + n\overline{\tau}\theta \\ \hat{\theta}_{\text{ML}} &= 1/\overline{\tau} \\ \mathbb{P}(\tau < 10) &= \int_0^{10} \hat{\theta}_{\text{ML}} e^{-\tau\hat{\theta}_{\text{ML}}} d\tau \\ &= 1 - e^{-10\hat{\theta}_{\text{ML}}}. \end{aligned}$$

4. A coin has Prob{heads} $= \theta$. To estimate θ, the following experiment is performed n times: the coin is flipped until 10 heads have been observed, and the total number of flips X is recorded. Each of the n experiments can be assumed to be independent. If the values $x_1, \ldots, x_n$ are observed, find the MLE of θ. (HINT: consider the negative binomial distribution.) Explain why your MLE result intuitively makes sense, e.g. by considering what happens when the true $\theta = 1/2$.

SOLUTION: Using the negative binomial distribution to describe the outcome of a single experiment, we have that

$$p(x_i|\theta) = \binom{x_i - 1}{9}\theta^{10}(1-\theta)^{x_i - 10}, x_i \geq 10,$$

therefore for the n independent experiments we have that

$$p(x|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \binom{x_i - 1}{9}\theta^{10}(1-\theta)^{x_i - 10} = \left[\prod_{i=1}^{n} \binom{x_i - 1}{9}\right]\theta^{10n}(1-\theta)^{s(x) - 10n},$$

where $s(x) = \sum_{i=1}^{n} x_i$. Now that we have an expression for the likelihood, we can go about finding the MLE. Instead of maximizing $p(x|\theta)$ directly, we can also maximize

$$\log p(x|\theta) = \log\left[\prod_{i=1}^{n} \binom{x_i - 1}{9}\right] + 10n \log\theta + (s(x) - 10n)\log(1-\theta).$$

Differentiating we find that

$$\frac{\partial}{\partial\theta}\log p(x|\theta) = \frac{10n}{\theta} - \frac{s(x) - 10n}{1-\theta}$$

which is equal to zero if

$$10n(1-\theta) = \theta(s(x) - 10n) \quad \Rightarrow \quad \hat{\theta}_{\text{MLE}} = \frac{10n}{s(x)} = \frac{10}{\frac{1}{n}\sum_{i=1}^{n} x_i}.$$

If we think about this estimator, it makes good sense. Lets say we had a fair coin in that $\theta = 1/2$, then on average we would need about 20 coin flips before we see 10 heads. This is reflected in our estimator, if the average outcome of our experiments is 20, then $\hat{\theta}_{\text{MLE}} = 1/2$.

5. Consider n iid observations from the exponential family of PDFs

$$p(x|\theta) = \exp[a(\theta)b(x) + c(x) + d(\theta)]$$

where a, b, c, d are functions of their respective arguments and x, θ are scalars.

a) Find an equation to be solved for the MLE.
SOLUTION:

$$
\begin{aligned}
p(x|\theta) &= \prod_{i=1}^{n} p(x_i|\theta) && \text{(1a)}\\
&= \prod_{i=1}^{n} \exp\{a(\theta)b(x_i) + c(x_i) + d(\theta)\} && \text{(1b)}\\
\log p(x|\theta) &= \sum_{i=1}^{n} \left[a(\theta)b(x_i) + c(x_i) + d(\theta)\right] && \text{(1c)}\\
\frac{d \log p(x|\theta)}{d\theta} &= \sum_{i=1}^{n} \left[\frac{da(\theta)}{d\theta} b(x_i) + \frac{dd(\theta)}{d\theta}\right] && \text{(1d)}\\
\text{solve } 0 &= \sum_{i=1}^{n} \left[a'(\theta)b(x_i) + d'(\theta)\right] && \text{(1e)}\\
&= a'(\theta)\sum_{i=1}^{n} b(x_i) + nd'(\theta) && \text{(1f)}
\end{aligned}
$$

b) Apply your result to the Poisson distribution.
SOLUTION:

$$
\begin{aligned}
p(x_i|\theta) &= e^{-\theta}\frac{\theta^{x_i}}{x_i!}\\
&= \exp\{\underbrace{\log\theta}_{a(\theta)}\underbrace{x_i}_{b(x_i)} + \underbrace{-\log(x_i!)}_{c(x_i)} + \underbrace{-\theta}_{d(\theta)}\}
\end{aligned}
$$

Solve (1f):

$$
\frac{1}{\theta}\sum_{i=1}^{n} x_i - n = 0 \Longrightarrow \hat{\theta}_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}
$$

c) Apply your result to the exponential distribution.
SOLUTION:

$$
\begin{aligned}
p(x_i|\theta) &= \theta e^{-\theta x_i}\\
&= \exp\{\underbrace{-\theta}_{a(\theta)}\underbrace{x_i}_{b(x_i)} + \underbrace{1}_{c(x_i)} + \underbrace{\log\theta}_{d(\theta)}\}
\end{aligned}
$$

Solve (1f):

$$
-\sum_{i=1}^{n} x_i + \frac{n}{\theta} = 0 \Longrightarrow \hat{\theta}_{\text{MLE}} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}
$$

6. Suppose $\boldsymbol{x} = [x_1, \ldots, x_n]^T$, where the x_i are independent and identically distributed according to $\mathcal{N}(\mu, \sigma^2)$. Consider the estimates of σ^2

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2,$$

where $\overline{x} = (1/n)\sum_{i=1}^{n} x_i$. Determine the bias and variance of each estimator. Which has the smaller MSE? (HINT: use the statistics (i.e. mean and variance) of chi-squared (χ^2) random variables. Cochran's Theorem may also be useful: `https://en.wikipedia.org/wiki/Cochran's_theorem`.)

SOLUTION: There are two ways of doing this problem.
Easy Way: Using the Cochran Theorem we know that $\bar{x} \sim N(\mu, \sigma^2/n)$ and $s^2 = \sigma^2/(n-1)\chi_{n-1}$, where χ_{n-1} is distributed according to a Chi-squared distribution with $n-1$ degrees of freedom. Note that $\mathbb{E}[\chi_k] = k$ and $\mathrm{Var}[\chi_k] = 2k$. From this we know that

$$\begin{aligned}
\mathbb{E}[s^2] &= \frac{\sigma^2}{n-1}\mathbb{E}[\chi_{n-1}] = \frac{\sigma^2}{n-1}(n-1) = \sigma^2, \\
\Rightarrow \mathrm{Bias}(s^2) &= \mathbb{E}[s^2] - \sigma^2 = 0, \\
\mathrm{Var}[s^2] &= \left(\frac{\sigma^2}{n-1}\right)^2 \mathrm{Var}[\chi_{n-1}] = \frac{\sigma^4}{(n-1)^2}2(n-1) = \frac{2\sigma^4}{n-1}, \\
\Rightarrow \mathrm{MSE}[s^2] &= \mathrm{Bias}^2[s^2] + \mathrm{Var}[s^2] = \frac{2\sigma^4}{n-1}.
\end{aligned}$$

Since we can show $\hat{\sigma}^2 = ((n-1)/n)s^2$, we have that

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^2] &= \frac{n-1}{n}\mathbb{E}[s^2] = \frac{n-1}{n}\sigma^2 = \sigma^2 - \sigma^2/n, \\
\Rightarrow \mathrm{Bias}^2(\hat{\sigma}^2) &= \sigma^2 - \sigma^2/n - \sigma^2 = -\sigma^2/n, \\
\mathrm{Var}[\hat{\sigma}^2] &= \left(\frac{n-1}{n}\right)^2 \mathrm{Var}[s^2] = \frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2}, \\
\Rightarrow \mathrm{MSE}(\hat{\sigma}^2) &= \left(-\frac{\sigma^2}{n}\right)^2 + \frac{2(n-1)\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}.
\end{aligned}$$

To determine which has the lower MSE, note that $n > n - 1/2$ and $n > n-1$, which implies that $2n^2 > 2(n-1/2)(n-1) = (2n-1)(n-1)$ so therefore

$$\mathrm{MSE}[s^2] = \frac{2\sigma^4}{n-1} > \frac{(2n-1)\sigma^4}{n^2} = \mathrm{MSE}[\hat{\sigma}^2],$$

therefore $\hat{\sigma}^2$ has the smaller MSE, even though $\hat{\sigma}^2$ is a biased estimator for σ^2.
Hard Way: If we start with $\bar{x}$, we know that since it is a linear combination of Gaussian

random variables, it is itself Gaussian distributed, so it is sufficient to find the mean and variance to know its distribution. So we need to find

$$\mathbb{E}[\bar{x}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[x_i] = \frac{1}{n}n\mu = \mu$$
$$\mathrm{Var}[\bar{x}] = \mathbb{E}[\bar{x}^2] - \mathbb{E}[\bar{x}]^2 = \mathbb{E}[\bar{x}^2] - \mu^2$$

where

$$\mathbb{E}[\bar{x}^2] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[x_i x_j] = \frac{1}{n^2}(n(\mu^2+\sigma^2) + (n^2-n)\mu^2) = \mu^2 + \sigma^2/n,$$

so therefore

$$\mathrm{Var}[\bar{x}] = \sigma^2/n.$$

Note that we use the fact that

$$\mathbb{E}[x_i x_j] = \begin{cases} \mu^2+\sigma^2 & \text{if } i = j, \\ \mu^2 & \text{if } i \neq j. \end{cases}$$

So we know that $\bar{x} \sim \mathcal{N}(\mu, \sigma^2/n)$. So now we're going to change variables to simplify the algebra a bit, so we define $y_i = x_i - \mu$ and $\bar{y} = \bar{x} - \mu$, so now we're working with $y_i \sim \mathcal{N}(0, \sigma^2)$ and $\bar{y} \sim \mathcal{N}(0, \sigma^2/n)$. Now we can compute the bias of the two estimators. Since (using our substitution),

$$\mathbb{E}[\hat{\sigma}^2] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]$$
$$\mathbb{E}[s^2] = \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right],$$

we only need to work out

$$\mathbb{E}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}y_i^2 - n\bar{y}^2\right] = \sum_{i=1}^{n}\mathbb{E}[y_i^2] - n\mathbb{E}[\bar{y}^2] = n\sigma^2 - n(\sigma^2/n) = (n-1)\sigma^2,$$

so we can show

$$\mathbb{E}[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2, \qquad \mathbb{E}[s^2] = \frac{n-1}{n-1}\sigma^2 = \sigma^2,$$
$$\mathrm{Bias}[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\sigma^2/n, \qquad \mathrm{Bias}[s^2] = \sigma^2 - \sigma^2 = 0.$$

So now we need to compute the variance. Again we exploit the fact that since

$$\mathrm{Var}(\hat{\sigma}^2) = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n}y_i^2 - n\bar{y}^2\right],$$
$$\mathrm{Var}(s^2) = \frac{1}{(n-1)^2}\mathrm{Var}\left[\sum_{i=1}^{n}y_i^2 - n\bar{y}^2\right],$$

we only need to find

$$\begin{aligned}
\mathrm{Var}\left[\sum_{i=1}^n y_i^2 - n\bar{y}^2\right] &= \mathbb{E}\left[\left(\sum_{i=1}^n y_i^2 - n\bar{y}^2 - (n-1)\sigma^2\right)^2\right] \\
&= \mathbb{E}\Bigg[\left(\sum_{i=1}^n y_i^2\right)^2 + \left(n\bar{y}^2\right)^2 + \left((n-1)\sigma^2\right)^2 \\
&\qquad - 2\left(\sum_{i=1}^n y_i^2\right)\left(n\bar{y}^2\right) - 2\left(\sum_{i=1}^n y_i^2\right)\left((n-1)\sigma^2\right) + 2\left(n\bar{y}^2\right)\left((n-1)\sigma^2\right)\Bigg] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^n y_i^2\right)^2\right] + n^2\mathbb{E}\left[\bar{y}^4\right] + (n-1)^2\sigma^4 - 2n\mathbb{E}\left[\bar{y}^2\sum_{i=1}^n y_i^2\right] \\
&\qquad - 2(n-1)\sigma^2\mathbb{E}\left[\sum_{i=1}^n y_i^2\right] + 2n(n-1)\sigma^2\mathbb{E}\left[\bar{y}^2\right].
\end{aligned}$$

So now we need to find each expectation term. In no particular order:

$$\begin{aligned}
\mathbb{E}\left[\bar{y}^2\right] &= \sigma^2/n \text{ (using the variance of } \bar{y}\text{)}, \\
\mathbb{E}\left[\sum_{i=1}^n y_i^2\right] &= n\sigma^2 \text{ (using the variance of } y_i\text{)}, \\
\mathbb{E}\left[\bar{y}^4\right] &= 3\left(\frac{\sigma^2}{n}\right)^2 = \frac{3}{n^2}\sigma^4 \text{ (using the variance of } \bar{y}\text{, and 4th moment of a Gaussian)}, \\
\mathbb{E}\left[\left(\sum_{i=1}^n y_i^2\right)^2\right] &= \sum_{i=1}^n\sum_{j=1}^n \mathbb{E}\left[y_i^2 y_j^2\right] = n(3\sigma^4) + (n^2-n)\sigma^4 = (n^2+2n)\sigma^4,
\end{aligned}$$

where we use

$$\mathbb{E}[y_i^2 y_j^2] = \begin{cases} 3\sigma^4 & \text{if } i = j, \\ \sigma^4 & \text{if } i \neq j. \end{cases}$$

The most difficult term to get is

$$\mathbb{E}\left[\bar{y}\sum_{i=1}^n y_i^2\right] = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \mathbb{E}\left[y_i^2 y_j y_k\right] = \frac{1}{n^2}\left(n(3\sigma^4) + n(n-1)\sigma^4\right) = \frac{n+2}{n}\sigma^4,$$

where we know

$$\mathbb{E}\left[y_i^2 y_j y_k\right] = \begin{cases} 3\sigma^4 & \text{if } i = j = k \text{ (occurs } n \text{ times)}, \\ \sigma^4 & \text{if } i \neq j = k \text{ (occurs } (n-1) \text{ times for each } i\text{, total of } n(n-1) \text{ times} \\ 0 & \text{otherwise.} \end{cases}$$

So now we need to put it all back together:

$$\begin{aligned}\mathrm{Var}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right] &= (n^2+2n)\sigma^4 + n^2\frac{3}{n^2}\sigma^4 + (n-1)^2\sigma^4 - 2n\frac{n+2}{n}\sigma^4 \\ &\qquad - 2(n-1)\sigma^2 n\sigma^2 + 2n(n-1)\sigma^2\frac{1}{n}\sigma^2 \\ &= 2(n-1)\sigma^4.\end{aligned}$$

So now we can show that

$$\begin{aligned}\mathrm{Var}(\hat{\sigma}^2) &= \frac{2(n-1)\sigma^4}{n^2} \\ \mathrm{Var}(s^2) &= \frac{2(n-1)\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{n-1},\end{aligned}$$

and the MSE of the estimators follows exactly as in the first way.