STATS 200: Homework 5

Due Thursday, November 10, at 5PM

1. **The geometric model.** Suppose $X_1, \ldots, X_n \overset{IID}{\sim}$ Geometric(p), where Geometric(p) is the geometric distribution on the positive integers $\{1, 2, 3, \ldots\}$ defined by the PMF

$$f(x|p) = p(1-p)^{x-1},$$

with a single parameter $p \in [0, 1]$. Compute the method-of-moments estimate of p, as well as the MLE of p. For large n, what approximately is the sampling distribution of the MLE? (You may use, without proof, the fact that the Geometric(p) distribution has mean $1/p$.)

2. **Fisher information in the normal model.** Let $X_1, \ldots, X_n \overset{IID}{\sim} \mathcal{N}(\mu, \sigma^2)$. We showed in class that the MLEs for μ and σ^2 are given by $\hat{\mu} = \bar{X}$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$.

(a) By computing the Fisher information matrix $I(\mu, \sigma^2)$, derive the approximate joint distribution of $\hat{\mu}$ and $\hat{\sigma}^2$ for large n. (Hint: Substitute $v = \sigma^2$ and treat v as the parameter rather than σ.)

(b) Suppose it is known that $\mu = 0$. Compute the MLE $\tilde{\sigma}^2$ in the one-parameter sub-model $\mathcal{N}(0, \sigma^2)$. The Fisher information matrix in part (a) has off-diagonal entries equal to 0—when $\mu = 0$ and n is large, what does this tell you about the standard error of $\tilde{\sigma}^2$ as compared to that of $\hat{\sigma}^2$?

3. **Necessity of regularity conditions.** Let $X_1, \ldots, X_n \overset{IID}{\sim}$ Uniform$(0, \theta)$ for a single parameter $\theta > 0$. (Uniform$(0, \theta)$ denotes the continuous uniform distribution on $(0, \theta)$, having PDF

$$f(x|\theta) = \frac{1}{\theta}\mathbb{1}\{0 \le X \le \theta\}.)$$

(a) Compute the MLE $\hat{\theta}$ of θ. (Hint: Note that the PDFs $f(x|\theta)$ do not have the same support for all $\theta > 0$, and they are also not differentiable with respect to θ—you will need to reason directly from the definition of the MLE.)

(b) If the true parameter is θ, explain why $\hat{\theta} \le \theta$ always, and hence why it cannot be true that $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution to $\mathcal{N}(0, v)$ for any $v > 0$.

4. **Generalized method-of-moments and the MLE.**
Consider a parametric model $\{f(x|\theta) : \theta \in \mathbb{R}\}$ of the form

$$f(x|\theta) = e^{\theta T(x) - A(\theta)} h(x), \tag{1}$$

where T, A, and h are known functions.

(a) Show that the Poisson(λ) model is of this form, upon reparametrizing by $\theta = \log \lambda$. What are the functions $T(x)$, $A(\theta)$, and $h(x)$?

(b) For any model of the form (1), differentiate the identity

$$1 = \int e^{\theta T(x) - A(\theta)} h(x) dx$$

with respect to θ on both sides, to obtain a formula for $\mathbb{E}_\theta[T(X)]$. ($\mathbb{E}_\theta$ denotes expectation when $X \sim f(x|\theta)$.) Verify that this formula is correct for the Poisson example in part (a).

(c) The **generalized method-of-moments** estimator is defined by the following procedure: For a fixed function $g(x)$, compute $\mathbb{E}_\theta[g(X)]$ in terms of θ, and take the estimate $\hat{\theta}$ to be the value of θ for which

$$\mathbb{E}_\theta[g(X)] = \frac{1}{n} \sum_{i=1}^{n} g(X_i).$$

(The method-of-methods estimator discussed in class is the special case of this procedure for $g(x) = x$.) Let $X_1, \ldots, X_n \overset{IID}{\sim} f(x|\theta)$, where $f(x|\theta)$ is of the form (1), and consider the generalized method-of-moments estimator using the function $g(x) = T(x)$. Show that this estimator is the same as the MLE. (You may assume that the MLE is the unique solution to the equation $0 = l'(\theta)$, where $l(\theta)$ is the log-likelihood.)

(d) Explain why the MLE and the method-of-moments estimator (the usual one defined in class with $g(x) = x$) are the same for the Poisson(λ) model.

5. **Computing the Gamma MLE.**
(a) Implement a function that takes as input a vector of data values `X`, performs the Newton-Raphson iterations to compute the MLEs $\hat{\alpha}$ and $\hat{\beta}$ in the Gamma(α, β) model, and outputs $\hat{\alpha}$ and $\hat{\beta}$. (You may use the form of the Newton-Raphson update equation derived in class.)

In R, this function may be defined as

```
gammaMLE <- function(X) {
  ...
  return(c(ahat,bhat))
}
```

where `...` should be filled in with the code to compute the MLE estimates `ahat` and `bhat` from `X`. You may terminate the Newton-Raphson iterations when $|\alpha^{(t+1)} - \alpha^{(t)}|$ is sufficiently small: For example, the high-level organization of the code to compute $\hat{\alpha}$ can be

```
a.prev = -Inf
a = # fill in code to initialize alpha^{(0)}
while (abs(a-a.prev) > 1e-12) {
  # fill in code to compute alpha^{(t+1)} from alpha^{(t)} = a
  # set a.new to be alpha^{(t+1)}
  a.prev = a
  a = a.new
}
ahat = a
```

The log-gamma function $\log \Gamma(\alpha)$, digamma function $\psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$, and its derivative $\psi'(\alpha)$ (called the trigamma function) are available in R as `lgamma(alpha)`, `digamma(alpha)`, and `trigamma(alpha)` respectively.

(b) For $n = 500$, use your function from part (a) to simulate the sampling distributions of $\hat{\alpha}$ and $\hat{\beta}$ computed from $X_1, \ldots, X_n \overset{IID}{\sim} \text{Gamma}(1, 2)$. Plot histograms of the values of $\hat{\alpha}$ and $\hat{\beta}$ across 5000 simulations, and report the simulated mean and variance of $\hat{\alpha}$ and $\hat{\beta}$ as well as the simulated covariance between $\hat{\alpha}$ and $\hat{\beta}$. Compute the inverse of the Fisher Information matrix $I(\alpha, \beta)$ at $\alpha = 1$ and $\beta = 2$—do your simulations support that $(\hat{\alpha}, \hat{\beta})$ is approximately distributed as $\mathcal{N}((1, 2), \frac{1}{n} I(1, 2)^{-1})$? (You may use the formula for the Fisher information matrix $I(\alpha, \beta)$ and/or its inverse derived in class.)

In R, you may simulate $X_1, \ldots, X_n \overset{IID}{\sim} \text{Gamma}(\alpha, \beta)$ using

```
X = rgamma(n, alpha, rate=beta)
```

The sample variance of a vector of values `X` is given by `var(X)`, and the sample covariance between two vectors of values `X` and `Y` (of the same length) is given by `cov(X,Y)`.