S&DS 242/542: Homework 6

Due Wednesday, March 5, at 1PM

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

$$f(x \mid p) = p(1-p)^{x-1}$$

with a single parameter $p \in [0, 1]$. You may use without proof that this distribution has mean $1/p$ and variance $(1-p)/p^2$.

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?

2. **The negative binomial model.** Suppose $X_1, \ldots, X_n \overset{IID}{\sim}$ NegBinom(r, p), where NegBinom(r, p) is the negative binomial distribution on $\{0, 1, 2, 3 \ldots\}$ defined by the PMF

$$f(x \mid p) = \binom{x+r-1}{x}(1-p)^r p^x.$$

Here $r > 0$ is a fixed and known positive integer, and $p \in [0, 1]$ is the unknown parameter. You may use without proof that this distribution has mean $pr/(1-p)$ and variance $pr/(1-p)^2$.

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?

3. **Generalized method-of-moments and the MLE.**
Consider a parametric model $f(x \mid \theta)$ with parameter $\theta \in \mathbb{R}$, whose PDF takes a form

$$f(x \mid \theta) = e^{\theta T(x) - A(\theta)} h(x) \text{ for } x \in \mathcal{X} \qquad (*)$$

where $\mathcal{X}$ is the range of possible data values.

(a) Show that the model Pareto($\theta, 1$) is of this form, where $\mathcal{X} = [1, \infty)$. What are the functions $T(x)$, $A(\theta)$, and $h(x)$ for this Pareto model?

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

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

with respect to θ on both sides, to obtain a formula for $\mathbb{E}_\theta[T(X)]$ in terms of $A(\theta)$. Verify that your formula is correct for the Pareto model in part (a).

[You may use $\frac{d}{d\theta}\int_{\mathcal{X}} e^{\theta T(x)-A(\theta)}h(x)dx = \int_{\mathcal{X}} \frac{d}{d\theta}[e^{\theta T(x)-A(\theta)}h(x)]dx$ without justifying this exchange of differentiation in θ and integration in x.]

(c) Let $X_1, \ldots, X_n \overset{IID}{\sim} f(x \mid \theta)$ where $f(x \mid \theta)$ is of the form $(*)$, and consider the generalized method-of-moments estimator $\hat{\theta}$ based on $T(x)$, i.e. $\hat{\theta}$ is the value of θ for which

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

If the MLE is the unique solution to the equation $0 = \ell_n'(\theta)$ where $\ell_n(\theta)$ is the log-likelihood, show that this generalized method-of-moments estimator is the same as the MLE.

Use this to explain why the generalized method-of-moments estimator based on $T(x) = \log x$ in the Pareto$(\theta, 1)$ model coincides with the MLE.

4. **Confidence intervals for a binomial proportion.**
Let $X_1, \ldots, X_n \overset{IID}{\sim}$ Bernoulli(p), and let $\hat{p} = \bar{X}$. We compare two different ways to construct a 95% confidence interval for p, both based on the Central Limit Theorem result

$$\sqrt{n}(\hat{p} - p) \to \mathcal{N}(0, p(1-p)). \qquad (**)$$

(a) Use the plugin estimate $\hat{p}(1-\hat{p})$ for the variance $p(1-p)$ to write down a 95% confidence interval for p. This is the approach discussed in Lecture 13.

(b) Instead of using this plugin estimate, note that equation $(**)$ implies, for large n,

$$\mathbb{P}\left[-\sqrt{p(1-p)}z^{(\alpha/2)} \le \sqrt{n}(\hat{p}-p) \le \sqrt{p(1-p)}z^{(\alpha/2)}\right] \approx 1 - \alpha.$$

Solve the two equations $\sqrt{n}(\hat{p} - p) = \pm\sqrt{p(1-p)}z^{(\alpha/2)}$ for p in terms of $\hat{p}$, to obtain a different 95% confidence interval for p.

(c) Perform a simulation study to determine the true probability that the confidence intervals in parts (a) and (b) cover p, for the 9 combinations of sample sizes $n = 10$, 40, 100 and true parameters $p = 0.1$, 0.3, 0.5. Report the simulated coverage probabilities in two tables. Which interval construction yields true coverage closer to 95% for small values of n?

[For each combination of n and p, it may be helpful to perform at least 100,000 simulations. In R, you may simulate $\hat{p}$ directly as `phat = rbinom(1,n,p)/n`.]