Midterm Exam: Statistical Inference

Due: Wed Mar 6 1:15pm

The number X of failures before the α^{th} success, in independent Bernoulli trials all with success probability $p = \theta^{-1}$, has a Negative Bionomial distribution, $X \sim \mathsf{NB}(\alpha, \theta^{-1})$. For $\alpha \in \mathbb{R}_+$, $\theta \in \Theta = [1, \infty)$ and $x \in \mathbb{N} = \{0, 1, ...\}$ the probability mass function $f(x \mid \alpha, \theta) \equiv \mathsf{P}[X = x \mid \alpha, \theta]$ is given by

$$f(x \mid \alpha, \theta) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)\; x!} \theta^{-\alpha} (1 - \theta^{-1})^x, \qquad x \in \mathbb{N}. \tag{1}$$

The mean and variance are $\mathsf{E}[X] = \alpha(\theta - 1)$ and $\mathsf{V}[X] = \alpha\theta(\theta - 1)$. Please answer the following questions about this $\mathsf{NB}(\alpha, \theta^{-1})$ distribution.

1. Let X be a single observation from $f(x \mid \alpha, \theta)$.

   a. If we treat $\theta \geq 1$ as fixed and known, is this an exponential family in $\alpha \in \mathbb{R}_+$? If so, write $f(x \mid \alpha) = f(x \mid \alpha, \theta)$ in standard form

   $$f(x \mid \alpha) = e^{\eta(\alpha) \cdot T(x) - B(\alpha)}\; h(x)$$

   for suitable $q \in \mathbb{N}$, $\eta(\alpha) \in \mathbb{R}^q$, $T(x) \in \mathbb{R}^q$, $B(\alpha) \in \mathbb{R}$, and $h(x) \geq 0$ (specify q, η, T, B, and h); if not, explain why (no proof needed).

2. Let X be a single observation from $f(x \mid \alpha, \theta)$.

   a. If we treat $\alpha > 0$ as fixed and known, is this an exponential family in $\theta \in \Theta$? If so, write $f(x \mid \theta) = f(x \mid \alpha, \theta)$ in standard form

   $$f(x \mid \theta) = e^{\eta(\theta) \cdot T(x) - B(\theta)}\; h(x)$$

   for suitable $q \in \mathbb{N}$, $\eta(\theta) \in \mathbb{R}^q$, $T(x) \in \mathbb{R}^q$, $B(\theta) \in \mathbb{R}$, and $h(x) \geq 0$. If not, explain why (no proof needed).

   b. Find the Fisher Information $I(\theta)$.

3. For fixed $\alpha > 0$, with $\theta \in \Theta$ unknown let $\vec{x}_n = (x_1, ..., x_n)$ be a simple random sample of size $n \in \mathbb{N}$ from $f(x \mid \theta) = f(x \mid \alpha, \theta)$.

   a. Find a sufficient (for $\theta \in \Theta$) statistic $S(\vec{x}_n)$. Verify that your S is sufficient. Find its mean and variance, as functions of θ.

b. Find the *observed* information $i(\theta, \vec{x}_n)$. Does it depend on $\vec{x}_n$ at all? If so, is it only as a function of your sufficient statistic S?

4. With a random sample $\vec{x}_n$ from $f(x \mid \theta)=f(x \mid \alpha, \theta)$ with known $\alpha > 0$,

   a. Find the maximum likelihood estimator $\hat{\theta}_n(\vec{x}_n)$.

   b. Evaluate or approximate the squared-error risk function

   $$R(\theta, \hat{\theta}_n) = \mathsf{E}\big[\, |\hat{\theta}_n - \theta|^2 \mid \theta \,\big].$$

   Is $\hat{\theta}_n$ consistent? Why? Plot $R(\theta, \hat{\theta}_n)$ for $\alpha = 3$ and $n = 4$.

   c. Find the bias $\beta(\theta) = \mathsf{E}[\hat{\theta}_n - \theta]$.

   d. Find the (absolute) efficiency

   $$\mathsf{Eff}(\hat{\theta}_n) = \Big\{ n\, I(\theta)\, R(\theta, \hat{\theta}_n) \Big\}^{-1}$$

   for each $n \in \mathbb{N}$ and $\theta \in \Theta$. Is $\hat{\theta}_n$ asymptotically efficient? Why?

   e. Evaluate the MLE for $n = 4$, $\alpha = 3$ with data sets $\vec{x}_4 = (25, 35, 22, 18)$ and $\vec{x}_4 = (0, 0, 0, 0)$.

5. Now let's adopt a Bayesian perspective, with fixed $\alpha > 0$.

   a. If θ has prior distribution given by

   $$\pi_{ab}(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{-a-b} (\theta - 1)^{b-1}, \qquad \theta \in \Theta = [1, \infty)$$

   for some $a, b > 0$, verify that for $p < a$ the prior p^{th} mean of θ is $\mathsf{E}_{ab}[\theta^p] = \frac{\Gamma(a+b)\Gamma(a-p)}{\Gamma(a)\Gamma(a+b-p)}$, and for $p = 1$ in particular, $\mathsf{E}_{ab}[\theta] = 1 + b/(a-1)$.

   b. For this prior distribution $\pi_{ab}(\theta)$, find the posterior distribution $\pi_{ab}(\theta \mid \vec{x}_n)$ for a sample $\vec{x}_n = (x_1, ..., x_n)$ of size n and the posterior mean (Bayes estimator) $\bar{\theta}_n^{ab}(\vec{x}_n) = \mathsf{E}_{ab}[\theta \mid \vec{x}_n]$.

   c. Does the MLE $\hat{\theta}_n$ coincide with $\bar{\theta}_n^{ab}(\vec{x}_n)$ for any $a, b > 0$, or limit as a, b tend to extremes?

   d. Find the squared-error risk function

   $$R(\theta, \bar{\theta}_n^{ab}) = \mathsf{E}_{ab}\big[\, |\bar{\theta}_n^{\pi} - \theta|^2 \mid \theta \,\big].$$

   Is $\bar{\theta}_n^{ab}$ consistent? Why? Plot $R(\theta, \bar{\theta}_4^{ab})$ for α=3, n=4, a=b=2.

e. Find the efficiency

$$\mathsf{Eff}(\bar\theta_n^{ab}) = \Big\{ n\, I(\theta)\, R(\theta, \bar\theta_n^{ab}) \Big\}^{-1}$$

for each $n \in \mathbb{N}$, $\theta \in \Theta$, $\alpha > 0$, and $a, b \geq 0$.
Is $\bar\theta_n^{ab}$ asymptotically efficient? Why?

6. With $\alpha > 0$ known, find the Jeffreys prior $\pi_J(\theta)$ and posterior $\pi_J(\theta \mid \vec{x}_n)$ distributions, and the posterior mean $\bar\theta_n^J(\vec{x}_n) = \mathsf{E}^J[\theta \mid \vec{x}_n]$.

   a. Does the Jeffreys posterior distribution coincide with that from any of the priors π_{ab} from Problem 5a. above? For which a, b?

7. In the past three problems you have computed three estimators for θ above ($\hat\theta_n$, $\bar\theta_n^{ab}$, $\bar\theta_n^J$) and their risk functions $R(\theta, \cdot)$. Let $n = 4$, $\alpha = 3$.

   a. For which θ is $\bar\theta^J$ better than $\hat\theta$? For which θ is $\bar\theta_n^{ab}$ better than $\hat\theta$, for $a = b = 2$?
   b. **Briefly**, how would you choose among these estimators for a particular problem? If any features of the problem would be relevant, identify them and explain.

8. With a random sample $\vec{x}_n$ from $f(x \mid \theta) = f(x \mid \alpha, \theta)$ with $\alpha > 0$ known,

   a. Does the MLE $\hat\theta_n$ have an asymptotically normal distribution? If so, find the mean $\mu_n(\theta)$ and variance $\sigma_n^2(\theta)$ of $\hat\theta_n$ and their maximum likelihood estimates $\hat\mu_n(\vec{x}_n)$ and $\hat\sigma_n^2(\vec{x}_n)$.
   b. Assuming asymptotic normality find an asymptotic 90% confidence interval based on these estimators, *i.e.*, find **statistics** $L_n(X)$, $R_n(X)$ satisfying $\mathsf{P}\big[\theta \in [L_n, R_n]\big] \approx 0.90$ for large n.
   c. Calculate the intervals for the two data sets $\vec{x}_4 = (25, 35, 22, 18)$ and $\vec{x}_4 = (0, 0, 0, 0)$, with $n = 4$, $\alpha = 3$. Are they reasonable?