Fall 2014

1. Let p, q be values in $[0,1]$ and $\alpha \in (0,1]$. Assume α and q known, and that p is an unknown parameter we would like to estimate. A coin is tossed n times, resulting in the sequence of zero one valued random variables $X_1, \ldots, X_n$. At each toss, independently of all other tosses, the coin has probability p of success with probability α, and probability q of success with probability $1-\alpha$.

   a) Write out the probability function of the observed sequence, and compute the maximum liklihood estimate $\hat{p}$ of p, when p is considered a parameter over all of $\mathbb{R}$. Verify that when $\alpha = 1$, one recovers the standard estimator of the unknown probability.
   Solution. Note that
   $$P(X_i = 1) = \begin{cases} p & \text{with probability } \alpha \\ q & \text{with probability } 1-\alpha \end{cases}.$$
   Hence,
   $$P(X_i = 1) = p\alpha + q(1-\alpha),$$
   and
   $$P(X_i = 0) = 1 - P(X_i = 1) = 1 - p\alpha - q(1-\alpha).$$
   Thus, the likelihood function is
   $$\mathcal{L}(p; \boldsymbol{x}) = (p\alpha + q(1-\alpha))^{\sum x_i} (1 - p\alpha - q(1-\alpha))^{n - \sum x_i},$$
   and so, the log-likelihood function is
   $$\log \mathcal{L}(p; \boldsymbol{x}) = \sum_{i=1}^{n} x_i \log(p\alpha + q(1-\alpha)) + \left(n - \sum_{i=1}^{n} x_i\right) \log(1 - p\alpha - q(1-\alpha)).$$
   The first derivative of the log-likelihood function is
   $$\frac{d}{dp} \log \mathcal{L} = \frac{\alpha \sum x_i}{p\alpha + q(1-\alpha)} - \frac{\alpha(n - \sum x_i)}{1 - p\alpha - q(1-\alpha)},$$
   and since the second derivative of the log-likelihood function,
   $$\frac{d^2}{dp^2} \log \mathcal{L} = -\frac{\alpha^2 \sum x_i}{(p\alpha + q(1-\alpha))^2} - \frac{\alpha^2(n - \sum x_i)}{(1 - p\alpha - q(1-\alpha))^2},$$
   is always negative, we see that
   $$\hat{p} = \boxed{\frac{1}{\alpha}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - q(1-\alpha)\right)}$$
   is the MLE of p. When $\alpha = 1$, the MLE becomes
   $$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$
   which is the MLE for the Bernoulli distribution with parameter p, which is what we expected.

b) Show $\hat{p}$ is unbiased, and calculate its variance.
Solution. Note that

$$\mathbb{E}\hat{p} = \frac{1}{\alpha}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i - q(1-\alpha)\right) = \frac{1}{\alpha}\left(p\alpha + q(1-\alpha) - q(1-\alpha)\right) = p,$$

and so, $\hat{p}$ is unbiased. As for the variance, since $X_i \sim \text{Bernoulli}(p\alpha + q(1-\alpha))$, we have that

$$\text{Var}(X_i) = (p\alpha + q(1-\alpha))(1 - (p\alpha + q(1-\alpha))).$$

Hence,

$$\text{Var}(\hat{p}) = \frac{1}{\alpha^2}\text{Var}\left(\bar{X}_n - q(1-\alpha)\right) = \frac{1}{n\alpha^2}(p\alpha + q(1-\alpha))(1 - (p\alpha + q(1-\alpha))).$$

□

c) Calculate the information bound for p, and determine if it is achieved by $\hat{p}$.
Solution. The Fisher information is

$$\begin{aligned}\mathcal{I}(p) &= -\mathbb{E}\left[-\frac{\alpha^2\sum x_i}{(p\alpha + q(1-\alpha))^2} - \frac{\alpha^2(n - \sum x_i)}{(1 - p\alpha - q(1-\alpha))^2}\right] \\ &= \frac{n\alpha^2}{p\alpha + q(1-\alpha)} + \frac{n\alpha^2}{1 - (p\alpha + q(1-\alpha))} \\ &= \frac{n\alpha^2}{(p\alpha + q(1-\alpha))(1 - (p\alpha + q(1-\alpha)))},\end{aligned}$$

and so, the Cramer-Rao lower bound is

$$\mathcal{I}^{-1}(p) = \frac{1}{n\alpha^2}(p\alpha + q(1-\alpha))(1 - (p\alpha + q(1-\alpha))),$$

which is achieved by $\hat{p}$. □

d) If one of the other parameters is unknown, can p still be estimated consistently?
Solution. We can still get an MLE for p, and since MLEs are consistent, why not?

2. Let $\boldsymbol{X} \in \mathbb{R}^n$ be distributed according to the density of mass function $p(\boldsymbol{x};\theta)$ for $\theta \in \Theta \subset \mathbb{R}^d$.

a) State the definition for $T(\boldsymbol{X})$ to be sufficient for θ.

Solution. $T(\boldsymbol{X})$ is sufficient for θ if

$$P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = T(\boldsymbol{x}))$$

is independent from θ.

b) Prove that if the (discrete) mass functions $p(\boldsymbol{x};\theta)$ can be factored as $h(\boldsymbol{x})g(T(\boldsymbol{x}),\theta)$ for some functions h and g, then $T(\boldsymbol{X})$ is sufficient for θ.

Solution. Define $A_{T(\boldsymbol{x})} = \{\boldsymbol{y} : T(\boldsymbol{y}) = T(\boldsymbol{x})\}$. Now,

$$\begin{aligned}
P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = T(\boldsymbol{x})) &= \frac{P(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = T(\boldsymbol{x}))}{P(T(\boldsymbol{X}) = T(\boldsymbol{x}))} \\
&= \frac{P(\boldsymbol{X} = \boldsymbol{x})}{P(T(\boldsymbol{X}) = T(\boldsymbol{x}))} \\
&= \frac{P(\boldsymbol{X} = \boldsymbol{x})}{\sum_{\boldsymbol{y} \in A_{T(\boldsymbol{x})}} P(\boldsymbol{X} = \boldsymbol{y})} \\
&= \frac{h(\boldsymbol{x})g(T(\boldsymbol{x}),\theta)}{\sum_{\boldsymbol{y} \in A_{T(\boldsymbol{x})}} h(\boldsymbol{y})g(T(\boldsymbol{y}),\theta)} \\
&= \frac{h(\boldsymbol{x})g(T(\boldsymbol{x}),\theta)}{g(T(\boldsymbol{x}),\theta)\sum_{\boldsymbol{y} \in A_{T(\boldsymbol{x})}} h(\boldsymbol{y})} \\
&= \frac{h(\boldsymbol{x})}{\sum_{\boldsymbol{y} \in A_{T(\boldsymbol{x})}} h(\boldsymbol{y})}
\end{aligned}$$

is independent of θ, and so, $T(\boldsymbol{X})$ is sufficient for θ. □

c) Let $X_1, \dots, X_n$ be independent with the Cauchy distribution $\mathcal{C}(\theta), \theta \in \mathbb{R}$ given by

$$p(x;\theta) = \frac{1}{\pi(1 + (x-\theta)^2)}.$$

Prove that the unordered sample $S = \{X_1, \dots, X_n\}$ can be determined from any $T(\boldsymbol{X})$ sufficient for θ. [HINT: Produce a polynomial from which S can be determined.]

Solution. Basically, we want to show that S is minimal sufficient, i.e. for all sufficient statistic $T(\boldsymbol{X})$, $S(\boldsymbol{X}) = g(T(\boldsymbol{X}))$. Recall that if

$$\frac{f(x \mid \theta)}{f(y \mid \theta)} \text{ is constant as a function of } \theta \iff S(x) = S(y),$$

then $S(X)$ is minimal sufficient.

The reverse direction is trivial. As for the forward direction, suppose that $f(x \mid \theta)/f(y \mid \theta) = c \neq 0$. Then,

$$\prod_{i=1}^{n} \left(1 + (x_i - \theta)^2\right) = c \prod_{j=1}^{n} \left(1 + (y_j - \theta)^2\right).$$

Note that the LHS is a polynomial of θ of degree $2n$ with roots $\theta = x_i \pm i$ and the RHS is a polynomial of θ of degreen $2n$ with roots $\theta = y_j \pm i$. Since the roots must be equal we have that there exists a permutation that can match the roots up, i.e. $S(X) = S(Y)$. Hence, S is minimal sufficient, and we are done. □