Fall 2008

1. Let $p \in (0,1)$ and $q = 1-p$.

   a) Show that
$$P(X=-1) = p \qquad \text{and} \qquad P(X=k) = q^2 p^k, \qquad k = 0, 1, \ldots$$
   defines a probability distribution for the random variable X.

   Solution. Since
$$\sum_{k=-1}^{\infty} P(X=k) = p + \sum_{k=0}^{\infty} (1-p)^2 p^k = p + (1-p)^2 \frac{1}{1-p} = p + (1-p) = 1,$$
   we see that the above is a probability distribution for X. □

   b) Given the single observation X, the statistic X is sufficient; is X also complete?

   Solution. Suppose that we define g such that $g(-1) = -g(1)(1-p)^2$, and 0 everywhere else. Then,
$$\mathbb{E}g(X) = pg(-1) + \sum_{k=0}^{\infty} g(k)(1-p)^2 p^k = pg(0) + p\left(g(1)(1-p)^2 + g(-1)\right) + \sum_{k=2}^{\infty} g(k) p^k (1-p)^2 = 0.$$
   However, $P\left(g(X) = 0\right) \neq 1$, and so, we see that X is not complete. □

   c) Determine all unbiased estimators of p, given one observation of X from the family above. [HINT: Consider $T(X) = \mathbf{1}(X=-1)$.]

   Solution. Suppose that $g(X)$ is an unbiased estimator of p. Then, we have
$$p = \mathbb{E}g(X) = pg(-1) + (1-p)^2 \sum_{k=0}^{\infty} g(k) p^k.$$
   Since
$$\begin{aligned} \frac{1}{(1-p)^2} = \left(\sum_{k=0}^{\infty} p^k\right)^2 &= \left(1 + p + p^2 + \cdots\right)\left(1 + p + p^2 + \cdots\right) \\ &= 1 + 2p + 3p^2 + 4p^3 + \cdots \\ &= \sum_{k=0}^{\infty} (k+1) p^k, \end{aligned}$$
   setting $\alpha = 1 - g(-1)$, we see that
$$\frac{\alpha p}{(1-p)^2} = \alpha \sum_{k=0}^{\infty} (k+1) p^{k+1} = \sum_{k=0}^{\infty} g(k) p^k,$$
   from which, it follows that $g(k) = \alpha k$ for $k = 0, 1, 2, \ldots$, i.e.
$$g(x) = \begin{cases} 1-\alpha & x = -1 \\ \alpha x & x = 0, 1, 2, \ldots \end{cases}.$$
   Note that when $\alpha = 0$, g is precisely $\mathbf{1}(X=-1)$. Hence, any unbiased estimator of p is of the form
$$g(X) = a\mathbf{1}(X=-1) + (1-a)X,$$
   where $a \in \mathbb{R}$. □

d) Find the UMVUE of p, or prove that it does not exist.

Solution. From part (c), we saw that all unbiased estimator of p are of the form

$$g_\alpha(X) = \alpha \mathbf{1}(X = -1) + (1 - \alpha)X.$$

Note that

$$\begin{aligned}\sum_{k=0}^{\infty} k^2 p^k &= \sum_{k=0}^{\infty} k(k-1)p^k + \sum_{k=0}^{\infty} kp^k = p^2 \sum_{k=0}^{\infty} k(k-1)p^{k-2} + p \sum_{k=0}^{\infty} kp^{k-1} \\ &= p^2 \left(\frac{1}{1-p}\right)'' + p\left(\frac{1}{1-p}\right)' = \frac{2p^2}{(1-p)^3} + \frac{p}{(1-p)^2}.\end{aligned}$$

Hence, the second moment of $g_\alpha(X)$ is

$$\begin{aligned}\mathbb{E}g_\alpha(X)^2 &= (1-\alpha)^2 P(X = -1) + \sum_{k=0}^{\infty} \alpha^2 k^2 P(X = k) \\ &= (1-\alpha)^2 p + \alpha^2 (1-p)^2 \sum_{k=0}^{\infty} k^2 p^k \\ &= (1-\alpha)^2 p + \alpha^2 (1-p)^2 \left[\frac{2p^2}{(1-p)^3} + \frac{p}{(1-p)^2}\right] \\ &= (1-\alpha)^2 p + \alpha^2 \frac{p+p^2}{1-p}.\end{aligned}$$

Thus, the variance of $g_\alpha(X)$ is

$$\mathbb{E}g_\alpha(X)^2 - (\mathbb{E}g_\alpha(X))^2 = \left(p + \frac{p+p^2}{1-p}\right)\alpha^2 - 2p\alpha + p - p^2 = \frac{2p}{1-p}\alpha^2 - 2p\alpha + (p - p^2),$$

which is a quadratic in α with positive coefficient. Taking the derivative and setting it equal to zero gives us that the minimal variance occurs at

$$\alpha = \frac{1-p}{2},$$

which depends on p. Hence, there does not exist a UMVUE of p. □

2. Consider the Pareto distribution $P(a,c)$, with positive parameters a and c, whose density function is given by

$$p(x;a,c) = \frac{ac^a}{x^{a+1}} \qquad \text{for } x \geq c.$$

a) Verify $p(x;a,c)$ is a density function, and find the associated distribution function.
Solution. Since

$$\int_c^\infty \frac{ac^a}{x^{a+1}}\,dx = -\left.\frac{c^a}{x^a}\right|_c^\infty = 1,$$

we see that $p(x;a,c)$ is indeed a density function. The associated distribution function is

$$F_X(x) = P(X \leq x) = \int_c^x \frac{ac^a}{t^{a+1}}\,dt = -\left.\left(\frac{c}{t}\right)^a\right|_c^x = 1 - \left(\frac{c}{x}\right)^a.$$

$\square$

b) When X has density $p(x;a,c)$, determine the distribution of $Y = \log X$. Note that

$$F_Y(y) = P(\log Y \leq y) = P(Y \leq e^y) = 1 - \frac{c^a}{e^{ay}},$$

and so, the density function of $Y = \log X$ is

$$f(y;a,c) = ac^a \exp\left(-ay\right), y \geq \log c,$$

which is the exponential distribution with parameter a shifted $\log c$. $\square$

c) Let $X_1, \ldots, X_n$ be a random sample from the Pareto $P(a,c)$ distribution. Find the maximum likelihood estimators $\hat{a}$ and $\hat{c}$ of a and c, respectively.
Solution. The likelihood function is

$$\mathcal{L}(a,c;\boldsymbol{x}) = \prod_{i=1}^n \frac{ac^a}{x_i^{a+1}} = a^n c^{an}\left(\prod_{i=1}^n x_i\right)^{-(a+1)},$$

and so, the log-likelihood function is

$$\log\mathcal{L}(a,c;\boldsymbol{x}) = n\log a + an\log c - (a+1)\sum_{i=1}^n \log x_i.$$

Taking the partial with respect to c, we get

$$\frac{\partial}{\partial c}\log\mathcal{L} = \frac{an}{c},$$

which is always increasing, and so, the log-likelihood reaches its maximum with respect to c when

$$\boxed{\hat{c} = \min X_i = X_{(1)}}.$$

Taking the partial of the log-likelihood with respect to a, we get

$$\frac{\partial}{\partial a}\log\mathcal{L} = \frac{n}{a} + n\log c - \sum_{i=1}^n \log x_i,$$

and since

$$\frac{\partial^2}{\partial a^2}\log\mathcal{L} = -\frac{n}{a^2} < 0,$$

we see that the MLE for a is

$$\boxed{\hat{a} = \frac{n}{\sum_{i=1}^n \log X_i - n\log\hat{c}} = \frac{n}{\sum_{i=1}^n \log\left(X_i/X_{(1)}\right)}}.$$

$\square$

d) Determine the distribution of $\hat{c}$ or of $2na/\hat{a}$.

Solution. The distribution function of $\hat{c}$ is

$$F_{\hat{c}}(x) = P(\min X_i \le x) = 1 - P(\min X_i \ge x) = 1 - \prod_{i=1}^{n} P(X_i \ge x) = 1 - \left(\frac{c}{x}\right)^{an},$$

and so, $\hat{c}$ has density

$$f(x; a, c) = an\frac{c^{an}}{x^{an+1}}$$

for $x \ge c$. In part (b), we saw that $\log X_i$ has an exponential distribution, and, with some manipulation, we can see that $2na/\hat{a}$ can be expressed as a sum of iid exponential distributions, which has the Gamma distribution. ☐