# MATH-562: Statistical Inference

Anthony Davison

**Problem 1** Let $X_1$ and $X_2$ be independent with distribution $\mathrm{P}(X = \theta - 1) = \mathrm{P}(X = \theta + 1) = 1/2$.

(a) Show that the set $\mathcal{C}$ that equals $\{(X_1 + X_2)/2\}$ when $X_1 \neq X_2$ and equals $\{X_1 - 1\}$ when $X_1 = X_2$ contains $\theta$ with probability $3/4$. Is $\mathcal{C}$ a sensible 75% confidence set?

(b) Sketch the sample space in this example and discuss possible reference sets. How would you construct 100% confidence sets for $\theta$?

**Problem 2**

(a) Compute the likelihood quantities for the exponential model $\phi \exp(-\phi y)$, for $y > 0$, expressed in terms of $\phi > 0$ and the mean $\theta = 1/\phi$, and verify that they transform as described on slide 63.

(b) A log-normal random variable is defined as $Y = e^X$, where $X \sim \mathcal{N}(\mu, \sigma^2)$. Given that $X$ has moment-generating function $M_X(t) = \exp(t\mu + t^2\sigma^2/2)$, show that

$$\mathrm{E}(Y) = \exp(\mu + \sigma^2/2) = \psi, \quad \mathrm{var}(Y) = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1) = \psi^2\lambda,$$

say and express $\mu$ and $\sigma^2$ in terms of $\psi$ and $\lambda$. Find the maximum likelihood estimates of $\psi$ and $\lambda$ based on a log-normal random sample $Y_1, \ldots, Y_n$.

**Problem 3** When independent positive continuous observations $Y_1, \ldots, Y_n$ with density function $f(y)$, survival function $\mathcal{F}(y) = \mathrm{P}(Y > y)$ and hazard function $h(y) = f(y)/\mathcal{F}(y)$ are right-censored at a constant $c$, the observed quantities are $(T_j, D_j) = (\min(Y_j, c), I(Y_j < c))$. This is *Type I censoring.*

(a) An old name for $h(t)$ is the *force of mortality.* Explain why, and show that the likelihood contribution based on $(t, d)$ can be written in the form $h(t)^d \mathcal{F}(t)$.

(b) When $Y_j \overset{\text{iid}}{\sim} \exp(\lambda)$, find the hazard and survival functions and hence show that the log likelihood can be written as $\sum_{j=1}^n (d_j \log\lambda - \lambda t_j)$. Find the maximum likelihood estimate of $\lambda$, and show that the expected information is $n(1 - e^{-\lambda c})/\lambda^2 = \imath(\lambda, c)$, say. Does this formula make sense?

(c) Show that if the censoring time $c$ is a realization of a random variable $C$ with gamma density $f(c) = (\lambda\alpha)^\nu c^{\nu-1}\exp(-c\lambda\alpha)/\Gamma(\nu)$, for $c > 0$ and $\alpha, \nu > 0$, then the expected information for $\lambda$ after averaging over $C$ is $\imath(\lambda) = n\{1 - (1 + 1/\alpha)^{-\nu}\}/\lambda^2$. Discuss the behaviour of $\imath(\lambda)$ when (i) $\alpha \to 0$, (ii) $\alpha \to \infty$, (iii) $\alpha = 1$, $\nu = 1$, (iv) $\alpha, \nu \to \infty$ with fixed $\mu = \nu/\alpha$.

*Hint:* $\mathrm{E}(C) = \nu/(\lambda\alpha)$ *and* $\mathrm{var}(C) = \mathrm{E}(C)^2/\nu$.

**Problem 4** In current status data all that is known about individuals is their status at a single time. For example, at time zero $n$ skiers are struck by an avalanche, and when rescuers locate skier $j$ at a later time $c_j$ they find that s/he is either alive (1) or dead (0).

(a) Show that the resulting likelihood can be written as $\prod_{j=1}^n F(c_j)^{1-d_j}\{1 - F(c_j)\}^{d_j}$. On what assumptions does this depend?

(b) If $F(x) = 1 - \exp(-\lambda x)$, for $\lambda > 0$ and $x > 0$, and all the $c_j$ are equal, then find the maximum likelihood estimator of $\lambda$ and the corresponding Fisher information.

(c) Find the asymptotic relative efficiency of the estimator in (b) relative to the maximum likelihood estimator when the observation is $(Y, D) = (\min(T, c), I(T > c))$, and $T \sim \exp(\lambda)$, i.e., the failure time is observed exactly up to time $c$, but is right-censored at $c$, and $D$ is the indicator of survival beyond $c$.