MATH562 – Fall 2025

Problem Set: Week 6

1. Let X_1 and X_2 be independent with distribution $\Pr(X = \theta - 1) = \Pr(X = \theta + 1) = 1/2$.

   (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 θ with probability 3/4. Is $\mathcal{C}$ a sensible 75% confidence set?

   Solution: If $X_1 \neq X_2$ then $(X_1 + X_2)/2 = (\theta + 1 + \theta - 1)/2 = \theta$, so in this case $\Pr(\theta \in \mathcal{C}) = 1$. If $X_1 = X_2$ then $\Pr(X_1 - 1 = \theta) = \Pr(X_1 = \theta + 1) = 1/2$. Moreover $\Pr(X_1 = X_2) = \Pr(X_1 = X_2 = \theta - 1) + \Pr(X_1 = X_2 = \theta + 1) = (1/2)^2 + (1/2)^2 = 1/2$, by independence, so

   $$\Pr(\theta \in \mathcal{C}) = \Pr(\theta \in \mathcal{C} \mid X_1 = X_2) \Pr(X_1 = X_2) + \Pr(\theta \in \mathcal{C} \mid X_1 \neq X_2) \Pr(X_1 \neq X_2) = 1/2 \times 1/2 + 1 \times 1/2 = 3/4,$$

   as required. This does not seem sensible, since if $X_1 \neq X_2$ we know θ with certainty, i.e., our confidence set has coverage 100%.

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

   Solution: The sample space consists of the three lines $\mathcal{Y} = \{(x, y) : |x - y| = 0, 2\}$. Two possible classes of reference sets are $\mathcal{S}_1(x) = \{(x, y) : |x - y| = 2\}$ and $\mathcal{S}_2(x) = \{(x, y) : x = y\}$. If $(x_1, x_2) \in \mathcal{S}_1(x)$ for some x, then $\mathcal{C}_1 = \{(x_1 + x_2)/2\}$ is a 100% confidence set, and if $(x_1, x_2) \in \mathcal{S}_2(x)$ for some x, then $\mathcal{C}_2 = \{x_1 + 1, x_1 - 1\}$ is a 100% confidence set.

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 4 (Likelihood slides).

   Solution: In this case

   $$\ell(\theta) = -y/\theta - \log\theta, \quad \tilde{\ell}(\phi) = \log\phi - \phi y, \quad \phi = 1/\theta, \quad \theta > 0,$$

   so $\partial\phi/\partial\theta = -1/\theta^2 = -\phi^2$, $\partial^2\phi/\partial\theta^2 = 2/\theta^3 = 2\phi^3$, and (in shorthand notation)

   $$\ell'(\theta) = y/\theta^2 - 1/\theta, \quad \ell''(\theta) = -2y/\theta^3 + 1/\theta^2, \quad \tilde{\ell}'(\phi) = 1/\phi - y, \quad \tilde{\ell}''(\phi) = -1/\phi^2.$$

   Note that $\hat{\phi} = 1/y$ and $\hat{\theta} = y$, so $\hat{\phi} = 1/\hat{\theta}$; checking the rest is tedious but easy.

   (b) A log-normal random variable is defined as $Y = e^X$, where $X \sim N(\mu, \sigma^2)$. 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 μ and σ^2 in terms of ψ and λ. Find the maximum likelihood estimates of ψ and λ based on a log-normal random sample $Y_1, \ldots, Y_n$.

   Solution: We need $\mathrm{E}(Y^r) = \mathrm{E}(e^{rX}) = M_X(r)$ for $r = 1, 2$, and using the given formula leads to

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

   as required. Hence $\sigma^2 = \log(1 + \lambda)$ and $\mu = \log\psi - \frac{1}{2}\log(1 + \lambda)$.

   The brute force approach to maximum likelihood estimation of ψ and λ would be to compute the density of Y, which is (check this if it is not obvious)

   $$f_Y(y; \mu, \sigma^2) = f_X(x; \mu, \sigma^2) \left|\frac{\partial x}{\partial y}\right|\Bigg|_{x = \log y} = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\log y - \mu)^2}{2\sigma^2}\right\}, \quad y > 0, \mu \in \mathbb{R}, \sigma^2 > 0,$$

then write $\sigma^2 = \log(1+\lambda)$ and $\mu = \log\psi - \frac{1}{2}\log(1+\lambda)$, compute the corresponding log likelihood based on $y_1, \ldots, y_n$ and hence obtain $\hat\psi$ and $\hat\lambda$. However it is much simpler to appeal to invariance: $\log Y_1, \ldots, \log Y_n i.i.d. N(\mu, \sigma^2)$, so

$$\hat\mu = n^{-1}\sum_{j=1}^n \log y_j, \quad \hat\sigma^2 = n^{-1}\sum_{j=1}^n (\log y_j - \hat\mu)^2,$$

and $\hat\psi = \exp(\hat\mu + \hat\sigma^2/2)$ and $\hat\lambda = \exp(\hat\sigma^2) - 1$.

3. When independent positive continuous observations $Y_1, \ldots, Y_n$ with density function $f(y)$, survival function $\mathcal{F}(y) = \Pr(Y > y)$ 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 $h(t)^d \mathcal{F}(t)$. **Solution**: The hazard function may be interpreted as $\lim_{h\to 0} h^{-1}\Pr\{Y \in [y, y+h) \mid Y > y\}$, i.e., the instantaneous probability of failure at time y conditional on survival to then, accounting for the term 'force of mortality'.

   Since $f(y) = h(y)\mathcal{F}(y)$ the likelihood contribution is $f(t)^d\mathcal{F}(t)^{1-d} = h(y)^d\mathcal{F}(t)$, as stated.

   (b) When Y_j *i.i.d.* $\exp(\lambda)$, 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 λ, and show that the expected information is $n(1-e^{-\lambda c})/\lambda^2 = \imath(\lambda, c)$, say. Does this formula make sense? **Solution**: In this case the observed data are $(t_1, d_1), \ldots, (t_n, d_n)$, $h(y) = \lambda$ and $\mathcal{F}(y) = \exp(-\lambda y)$, so the log likelihood is

$$\ell(\lambda) = \sum_{j=1}^n \{d_j \log h(t_j) + \log\mathcal{F}(t_j)\} = \sum_{j=1}^n (d_j\log\lambda - \lambda t_j), \quad \lambda > 0.$$

   Differentiation gives $\hat\lambda = \sum d_j / \sum t_j$ and $\ell''(\lambda) = -\sum d_j/\lambda^2$ and as $D_j = I(Y_j \le c)$ is an indicator variable, $\mathrm{E}(D_j) = \Pr(Y_j \le c) = 1 - e^{-\lambda c}$, leading to the stated formula. This seems reasonable because a proportion $e^{-\lambda c}$ of the data are lost to censoring, and the lack of memory property means that if they exceed c we have no idea of their values and thus no information on λ from them.

   (c) In (b) show that if the censoring time c is viewed as 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 λ 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* $\mathrm{Var}(C) = \mathrm{E}(C)^2/\nu$. **Solution**: Treating c as a realisation of C means that we we regard the expected information computed in (b) as conditional on $C = c$, so the unconditional Fisher information is $\imath(\lambda) = \mathrm{E}_C\{\imath(\lambda, C)\}$. To compute this we need $\mathrm{E}_C(e^{-\lambda C})$, which is the moment-generating function $M_C(t)$ of C evaluated at $t = -\lambda$. Hence

$$\imath(\lambda) = \mathrm{E}_C\{\imath(\lambda, C)\} = \frac{n}{\lambda^2}\left\{1 - \mathrm{E}\left(e^{-\lambda C}\right)\right\} = \frac{n}{\lambda^2}\left\{1 - \left(\frac{\lambda\alpha}{\lambda\alpha + \lambda}\right)^\nu\right\} = \frac{n}{\lambda^2}\left\{1 - (1 + 1/\alpha)^{-\nu}\right\}.$$

   (i) When $\alpha \to 0$, $\mathrm{E}(C) \to \infty$, so no observations will be censored in the limit, and thus $\imath(\lambda)$ tends to the usual quantity without censoring.

   (ii) When $\alpha \to \infty$, $\mathrm{E}(C) \to 0$, the censoring probability tends to unity, and thus $\imath(\lambda)$ tends to zero.

   (iii) When $\alpha = \nu = 1$ then $(1 + 1/\alpha)^{-\nu} = 1/2$, and $\imath(\lambda)$ is half that for an uncensored sample.

   (iv) When $\alpha, \nu \to \infty$ for fixed $\mu = \nu/\alpha$, $(1 + 1/\alpha)^{-\nu} = (1 + \mu/\nu)^{-\nu} \to e^{-\mu}$, and $\imath(\lambda)$ corresponds to censoring at a fixed time μ/λ; as $\mathrm{E}(C) = \nu/(\alpha\lambda) \to \mu/\lambda$ and $\mathrm{Var}(C) = \nu/(\alpha\lambda)^2 \to 0$, $C \xrightarrow{p} \mu/\lambda$.

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? **Solution**: If we assume that the times to death have common distribution F, then the probability of death by time c is $F(c)$, and the probability of being alive is thus $1-F(c)$. Hence if d is the indicator that the individual is alive, then their likelihood contribution is $F(c)^{1-d}\{1-F(c)\}^d$, which yields the given likelihood, if the outcomes are independent.

   (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 λ and the corresponding Fisher information. **Solution**: Writing $p(\lambda) = \exp(-\lambda c)$ and with $s = \sum_j d_j$ survivors, the log likelihood can be written as

$$\ell(\lambda) = (n-s)\log\{1-p(\lambda)\} + s\log p(\lambda), \quad \lambda > 0,$$

   so $p(\hat\lambda) = s/n$, which yields $\hat\lambda = c^{-1}\log(n/s)$. For the Fisher information we note that $S \sim B\{n, p(\lambda)\}$, and then after a little work obtain

$$\mathrm{E}\left\{-\frac{\partial^2\ell(\lambda)}{\partial\lambda^2}\right\} = -\,\mathrm{E}\left\{\left(\frac{\partial p}{\partial\lambda}\right)^2\frac{\partial^2\ell}{\partial p^2} + \frac{\partial^2 p}{\partial\lambda^2}\frac{\partial\ell}{\partial p}\right\} = \frac{nc^2p(\lambda)}{1-p(\lambda)},$$

   because $\partial p(\lambda)/\partial\lambda = -cp(\lambda)$, $\mathrm{E}(\partial\ell/\partial p) = 0$ and $\mathrm{E}(S) = np(\lambda)$.

   (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.

   Solution: In this case the likelihood contribution for an individual is $(\lambda e^{-\lambda y})^{1-d}(e^{-\lambda c})^d$, so with $s = \sum_j d_j$ the overall log likelihood is

$$\sum_{j=1}^n (1-d_j)(\log\lambda - \lambda y_j) - \lambda cs, \quad \lambda > 0.$$

   This has second derivative $-(n-s)/\lambda^2$, leading to Fisher information $\{n-\mathrm{E}(S)\}/\lambda^2 = n\{1-p(\lambda)\}/\lambda^2$ so the asymptotic relative efficiency of using current status data is

$$\frac{nc^2p(\lambda)}{1-p(\lambda)} \div \frac{n\{1-p(\lambda)\}}{\lambda^2} = \frac{\lambda^2c^2p(\lambda)}{\{1-p(\lambda)\}^2} = \frac{p(\lambda)\{\log p(\lambda)\}^2}{\{1-p(\lambda)\}^2}.$$

   Perhaps surprisingly, this is fairly high: it equals $0.999, 0.961, 0.655$ when $p = 0.9, 0.5, 0.1$ respectively, so despite the strong censoring, relatively little information is lost overall. Unfortunately this is a feature of the exponential distribution, and not of the problem itself.