STAT 460/560: Statistical Inference 1 **Winter Term 1 (Fall) 2026**

Assignment 5

Due date: Oct. 19, 11 p.m. *Instructor: Daniel McDonald and Benjamin Bloem-Reddy*

Instructions

Academic integrity policy: I encourage you to discuss verbally with other students about the assignment. However, you **must** write your answers by yourself. If you discuss the assignment with anyone (a classmate or anyone else), you must say so at the top of your solutions. Use of AI tools to produce full or partial answers to these questions or to related questions is strictly prohibited. This includes tools like Co-pilot, auto-complete, etc. Also, refrain from looking at answer keys from other schools, previous years, or the Math Stack Exchange. For more information, see: the syllabus or contact the instructor.

1. Solutions may be handwritten or typeset with LaTeX (see the course website for a template). It must be legible. Difficult-to-read solutions will be marked down or ignored.
2. *Justify your answers formally. (Err on the side of too much detail so that the teaching team knows exactly what you're doing and why.)*
3. Submit your work by 11:59 pm (Vancouver time) on the due date to Canvas (.pdf).
4. For questions that require computer simulations, you may use any language from the list of `(R, Julia, Python)`, with level of support provided by the teaching team decreasing further in the list. Please include in your PDF submission the relevant plots generated by your simulation code, and append your code (in PDF) to your submission. Please retain the code you used, as you may be asked for it if required.

Questions

1. Wasserman (2004) 9.6.

   Solution:

   Since $X_i \sim \mathsf{N}(\theta, 1)$, we have

   $$\psi = \mathbb{P}(Y = 1) = \mathbb{P}(X > 0) = \mathbb{P}(X - \theta > -\theta) = 1 - \Phi(-\theta) = \Phi(\theta),$$

   where Φ is the standard normal CDF. The parameter is θ, and $\psi = g(\theta) = \Phi(\theta)$ is a one-to-one function of θ.

   (a) The log-likelihood for θ is

   $$\ell(\theta) \propto -\tfrac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2,$$

   so $\hat{\theta} = \overline{X}_n$. By equivariance of the MLE,

   $$\hat{\psi} = \Phi(\hat{\theta}) = \Phi(\overline{X}_n).$$

(b) The Fisher information for one observation is $I(\theta) = 1$, because we know that the second derivative of the log-likelihood with respect to θ is -1. Therefore, $\mathrm{se}(\hat{\theta}) = 1/\sqrt{n}$. Since $g'(\theta) = \phi(\theta)$, the standard normal density, the delta method gives

$$\widehat{\mathrm{se}}(\hat{\psi}) = |g'(\hat{\theta})|\ \mathrm{se}(\hat{\theta}) = \frac{\phi(\overline{X}_n)}{\sqrt{n}},$$

and

$$\sqrt{n}\,(\hat{\psi} - \psi) \rightsquigarrow \mathsf{N}\left(0,\ \phi(\theta)^2\right).$$

An approximate 95% confidence interval is therefore

$$\hat{\psi} \pm 1.96\,\frac{\phi(\overline{X}_n)}{\sqrt{n}} = \Phi(\overline{X}_n) \pm 1.96\,\frac{\phi(\overline{X}_n)}{\sqrt{n}}.$$

(c) The $Y_i = \mathbf{I}(X_i > 0)$ are iid Bernoulli(ψ) with $\mathbb{E}[Y_i] = \psi$. By the weak law of large numbers,

$$\tilde{\psi} = \overline{Y}_n \xrightarrow{\mathbb{P}} \mathbb{E}[Y_1] = \psi,$$

so $\tilde{\psi}$ is consistent for ψ.

(d) From (b), the asymptotic standard deviation of $\hat{\psi}$ is $\phi(\theta)/\sqrt{n}$. For $\tilde{\psi}$, the standard deviation is

$$\mathrm{se}(\tilde{\psi}) = \sqrt{\frac{\psi(1-\psi)}{n}} = \sqrt{\frac{\Phi(\theta)\,(1-\Phi(\theta))}{n}}.$$

The asymptotic relative efficiency of $\tilde{\psi}$ to $\hat{\psi}$ is the ratio of asymptotic variances (MLE over competitor):

$$\mathrm{ARE}(\tilde{\psi}, \hat{\psi}) = \frac{\mathrm{AsyVar}(\hat{\psi})}{\mathrm{AsyVar}(\tilde{\psi})} = \frac{\phi(\theta)^2}{\Phi(\theta)\left(1 - \Phi(\theta)\right)}.$$

This is at most 1, as it must be since the MLE is asymptotically efficient. For example, at $\theta = 0$ we have $\phi(0)^2 = 1/(2\pi)$ and $\Phi(0)(1 - \Phi(0)) = 1/4$, so

$$\mathrm{ARE} = \frac{2}{\pi} \approx 0.637.$$

The ARE decreases toward 0 as $|\theta| \to \infty$, because the Gaussian density in the numerator decays much faster than $\Phi(1 - \Phi)$ in the denominator.

(e) Suppose the X_i are iid from some distribution F that is not $\mathsf{N}(\theta, 1)$, with $\mathbb{E}|X_1| < \infty$ and mean $\mu = \mathbb{E}[X_1]$. By the law of large numbers, $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$. Since Φ is continuous, the continuous mapping theorem gives

$$\hat{\psi} = \Phi(\overline{X}_n) \xrightarrow{\mathbb{P}} \Phi(\mu).$$

The true value of the parameter is $\psi = \mathbb{P}(X > 0) = 1 - F(0)$. Hence $\hat{\psi}$ is consistent only if

$$\Phi(\mu) = 1 - F(0),$$

which fails in general. For instance, if $X \sim \mathsf{N}(0, \sigma^2)$ with $\sigma \neq 1$, then $\psi = 1/2$, so this particular example is consistent; but if X is, say, exponential shifted to have mean 0, i.e. $X = E - 1$ with $E \sim \mathsf{Exp}(1)$, then $\mu = 0$, $\Phi(\mu) = 1/2$, and

$$\psi = \mathbb{P}(E > 1) = e^{-1} \approx 0.368 \neq 0.5,$$

so $\hat{\psi}$ is not consistent.

In general, $\hat{\psi}$ converges in probability to $\Phi(\mathbb{E}[X_1])$, which is the value of ψ *under the best-fitting normal model* (the KL-projection of F onto $\{\mathsf{N}(\theta, 1)\}$ has $\theta^* = \mathbb{E}[X_1]$), but this differs from the true $\psi = \mathbb{P}(X_1 > 0)$ unless F happens to satisfy $\Phi(\mu) = 1 - F(0)$. By contrast, $\tilde{\psi} = \overline{Y}_n \to \mathbb{P}(X_1 > 0)$ regardless of the distribution, so it is consistent under any F.

2. Wasserman (2004) 9.7 (skip d).

Solution:

(a) The log-likelihood is

$$\ell(p_1, p_2) = X_1 \log p_1 + (n_1 - X_1)\log(1 - p_1) + X_2 \log p_2 + (n_2 - X_2)\log(1 - p_2) + \text{const.}$$

It separates in p_1 and p_2, so each parameter can be maximized separately. Setting $\partial\ell/\partial p_j = X_j/p_j - (n_j - X_j)/(1 - p_j) = 0$ gives

$$\hat{p}_1 = \frac{X_1}{n_1}, \qquad \hat{p}_2 = \frac{X_2}{n_2}.$$

By equivariance of the MLE,

$$\hat{\psi} = \hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}.$$

(b) The second derivatives of the log-likelihood are

$$\frac{\partial^2 \ell}{\partial p_1^2} = -\frac{X_1}{p_1^2} - \frac{n_1 - X_1}{(1 - p_1)^2}, \qquad \frac{\partial^2 \ell}{\partial p_2^2} = -\frac{X_2}{p_2^2} - \frac{n_2 - X_2}{(1 - p_2)^2},$$

and the cross derivative is $\partial^2\ell/\partial p_1 \partial p_2 = 0$. Taking expectations with $\mathbb{E}[X_j] = n_j p_j$,

$$-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial p_j^2}\right] = \frac{n_j}{p_j} + \frac{n_j}{1 - p_j} = \frac{n_j}{p_j(1 - p_j)}.$$

Hence

$$I_n(p_1, p_2) = \begin{pmatrix} \dfrac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \dfrac{n_2}{p_2(1-p_2)} \end{pmatrix}, \qquad I_n^{-1}(p_1, p_2) = \begin{pmatrix} \dfrac{p_1(1-p_1)}{n_1} & 0 \\ 0 & \dfrac{p_2(1-p_2)}{n_2} \end{pmatrix}.$$

(c) The gradient of $g(p_1, p_2) = p_1 - p_2$ is

$$\nabla g = \begin{pmatrix} \partial g/\partial p_1 \\ \partial g/\partial p_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The delta method gives

$$\text{se}(\hat{\psi})^2 = (\nabla g)^\top \, I_n^{-1}(p_1, p_2) \, (\nabla g) = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}.$$

Therefore

$$\text{se}(\hat{\psi}) = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}},$$

and the estimated standard error, obtained by plugging in the MLEs, is

$$\widehat{\text{se}}(\hat{\psi}) = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}.$$

This agrees with the direct computation $\text{Var}(\hat{p}_1 - \hat{p}_2) = \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2)$, which holds exactly by independence. An approximate 95% confidence interval for ψ is then $\hat{\psi} \pm 1.96\, \widehat{\text{se}}(\hat{\psi})$.

3. Let p_θ the density on $\mathbb{R}^2$ that is uniform on the disk of radius θ centered at the origin. Let $X_1, \ldots, X_n$ be i.i.d. from p_θ. Write down the likelihood function for θ. Find a minimal sufficient statistic for θ. Show that X_1 is not sufficient for θ.

   Solution:

   The density of X_i is given by

   $$p_\theta(x) = \frac{1}{\pi\theta^2}\mathbf{I}\{\|x\| \le \theta\}.$$

   The likelihood function is

   $$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^n p_\theta(x_i) = \left(\frac{1}{\pi\theta^2}\right)^n \prod_{i=1}^n \mathbf{I}\{\|x_i\| \le \theta\} = \left(\frac{1}{\pi\theta^2}\right)^n \mathbf{I}\left\{\max_i \|x_i\| \le \theta\right\}.$$

   The likelihood is non-zero if and only if $\theta \ge \max_i \|x_i\|$. Therefore, the likelihood function can be rewritten as

   $$L(\theta; x_1, \ldots, x_n) = \left(\frac{1}{\pi\theta^2}\right)^n \mathbf{I}\{\theta \ge \max_i \|x_i\|\}.$$

   The minimal sufficient statistic is $T(X_1, \ldots, X_n) = \max_i \|X_i\|$. To show that X_1 is not sufficient, we can use the factorization theorem. The likelihood function cannot be factored into a product of a function of X_1 and a function of θ alone, since the likelihood depends on all X_i through the maximum. Therefore, X_1 is not sufficient for θ.

4. Let $X \sim \mathsf{Gamma}(\alpha, \beta)$. Show that this is an exponential family by expressing the density in the form $p_\theta(x) = h(x)\exp(\eta(\theta)^\top T(x) - A(\theta))$ for appropriate functions h, η, T, and A. Find a minimal sufficient statistic for θ. Finally, use Wasserman (2004, Theorem 9.47) to find the mean and variance of X.

   Solution:

   The density of X is given by

   $$\begin{aligned} p_\theta(x) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\mathbf{I}\{x > 0\} \\ &= \mathbf{I}\{x > 0\}\frac{1}{x}\exp\left(\frac{1}{\beta}(-x) + \alpha\log x + \alpha\log(1/\beta) - \log\Gamma(\alpha)\right). \end{aligned}$$

   So this is in (natural) exponential family form with

   - $h(x) = \mathbf{I}\{x > 0\}\frac{1}{x}$,
   - $\eta(\theta) = (\alpha, \frac{1}{\beta}) = (\eta_1, \eta_2)$,
   - $T(x) = (\log x, -x)$,
   - $A(\eta) = \log\Gamma(\eta_1) - \eta_1\log\eta_2$.

   Therefore, the Gamma distribution is indeed an exponential family.

   By the factorization theorem, a minimal sufficient statistic for θ is given by $T(X) = (\log X, X)$.

   Finally, using Wasserman (2004, Theorem 9.47), we can find the mean and variance of X.

   $$\mathbb{E}[X] = -\mathbb{E}[t_2(X)] = -\frac{\partial}{\partial\eta_2}A(\eta) = -\frac{\partial}{\partial\eta_2}(\log\Gamma(\eta_1) - \eta_1\log\eta_2) = \eta_1/\eta_2 = \alpha\beta,$$

   $$\mathrm{Var}(X) = \mathrm{Var}(-X) = \mathrm{Var}(t_2(X)) = \frac{\partial^2}{\partial\eta_2^2}A(\eta) = \frac{\partial^2}{\partial\eta_2^2}(\log\Gamma(\eta_1) - \eta_1\log\eta_2) = \eta_1/\eta_2^2 = \alpha\beta^2.$$

5. Wasserman (2004) 11.6.

   Solution:

(a) We have that the posterior is proportional to the likelihood times the prior, i.e.

$$\begin{aligned} p(\lambda \mid X_1, \ldots, X_n) &\propto \prod_{i=1}^{n} \lambda^{X_i} e^{-\lambda} \cdot \lambda^{\alpha-1} e^{-\lambda/\beta} \\ &= \lambda^{\sum_{i=1}^{n} X_i + \alpha - 1} e^{-(n+1/\beta)\lambda} \\ &= \lambda^{n\overline{X}+\alpha-1} e^{-(n+1/\beta)\lambda}. \end{aligned}$$

This is the kernel of a Gamma distribution with shape parameter $n\overline{X} + \alpha$ and rate parameter $n + 1/\beta$. The posterior mean is then

$$\mathbb{E}[\lambda \mid X_1, \ldots, X_n] = \frac{n\overline{X} + \alpha}{n + 1/\beta}.$$

(b) The Jeffreys' prior is given by $\pi(\lambda) \propto \sqrt{I(\lambda)}$, where $I(\lambda)$ is the Fisher information. The log-density is

$$\log p_\lambda(x) = x \log \lambda - \lambda - \log(x!),$$

with second derivative

$$\frac{\partial^2}{\partial \lambda^2} \log p_\lambda(x) = \frac{-x}{\lambda^2}.$$

The Fisher information is then

$$I(\lambda) = -\mathbb{E}\left[\frac{\partial^2}{\partial \lambda^2} \log p_\lambda(X)\right] = -\mathbb{E}\left[\frac{-X}{\lambda^2}\right] = \frac{\mathbb{E}[X]}{\lambda^2} = \frac{1}{\lambda}.$$

Therefore, the Jeffreys' prior is

$$\pi(\lambda) \propto \sqrt{\frac{1}{\lambda}} = \lambda^{-1/2}.$$

The posterior is then proportional to

$$\begin{aligned} p(\lambda \mid X_1, \ldots, X_n) &\propto \prod_{i=1}^{n} \lambda^{X_i} e^{-\lambda} \cdot \lambda^{-1/2} \\ &= \lambda^{\sum_{i=1}^{n} X_i - 1/2} e^{-n\lambda} \\ &= \lambda^{n\overline{X}-1/2} e^{-n\lambda}. \end{aligned}$$

This is the kernel of a Gamma distribution with shape parameter $n\overline{X} + 1/2$ and rate parameter n. The posterior mean is then

$$\mathbb{E}[\lambda \mid X_1, \ldots, X_n] = \frac{n\overline{X} + 1/2}{n}.$$

References

Larry Wasserman. *All of Statistics*. Springer, New York, 2004.