STA360/601 Midterm Solutions

1. (15 points)

   (a) (5 points) What is the formula for Bayes' theorem?

   $$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \quad \text{OR} \quad p(\theta|x) \propto p(x|\theta)p(\theta) \quad \text{OR} \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

   (b) (5 points) You receive x_i telephone calls on day i, for $i = 1, \ldots, n$. You wish to model this as $X_1, \ldots, X_n$ i.i.d. from some distribution. Which of the following distributions would make sense to use? (Circle one.)

      i. ~~Beta~~ (continuous, limited to $(0, 1)$)

      ii. [Poisson] (discrete, on $\{0, 1, 2, 3, \ldots\}$) (also see "law of small numbers")

      iii. ~~Bernoulli~~ (discrete, but limited to $\{0, 1\}$)

      iv. ~~Exponential~~ (continuous)

   (c) (5 points) Suppose $X, X_1, \ldots, X_N$ are i.i.d. and assume $\mathbb{E}|X| < \infty$ and $\mathbb{V}(X) < \infty$. What is the standard deviation of

   $$\frac{1}{N}\sum_{i=1}^{N} X_i?$$

   Hint: It is the same as the RMSE of the Monte Carlo approximation. (Circle one.)

      i. $\mathbb{V}(X)/N$

      ii. $\mathbb{V}(X)/\sqrt{N}$

      iii. $\sigma(X)/N$

      iv. [$\sigma(X)/\sqrt{N}$]

   $$\mathbb{V}\Big(\tfrac{1}{N}\sum X_i\Big) = \tfrac{1}{N^2}\mathbb{V}\big(\sum X_i\big) = \tfrac{1}{N^2}\sum_{i=1}^{N}\mathbb{V}\big(X_i\big) = \tfrac{1}{N}\mathbb{V}\big(X\big)$$

   $$\sigma\Big(\tfrac{1}{N}\sum X_i\Big) = \mathbb{V}\Big(\tfrac{1}{N}\sum X_i\Big)^{1/2} = \Big(\tfrac{1}{N}\mathbb{V}\big(X\big)\Big)^{1/2} = \tfrac{1}{\sqrt{N}}\sigma(X)$$

2. (17 points) (Marginal likelihood)
Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Geometric}(\theta)$ given θ. Consider a Beta(a, b) prior on θ. What is the marginal likelihood $p(x_{1:n})$?

(Your answer must be an explicit expression in terms of $a, b, x_1, \ldots, x_n, n$, and any of the special functions on page 2. You must show your work to receive full credit.)

$$\begin{aligned} p(x_i|\theta) &= \theta(1-\theta)^{x_i}\mathbb{1}(x_i \in \{0, 1, 2, \ldots\}) \\ p(\theta) &= \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}(0 < \theta < 1) \end{aligned}$$

For $x_1, \ldots, x_n \in \{0, 1, 2, \ldots\}$,

$$\begin{aligned} p(x_{1:n}) &= \int p(x_{1:n}|\theta)p(\theta)d\theta \\ &= \int \Big(\prod_{i=1}^{n} p(x_i|\theta)\Big)p(\theta)d\theta \\ &= \int \theta^n(1-\theta)^{\sum x_i}\frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}(0 < \theta < 1) \\ &= \frac{1}{B(a,b)}\int \theta^{a+n-1}(1-\theta)^{b+\sum x_i - 1}\mathbb{1}(0 < \theta < 1) \\ &= \frac{B(a+n,\, b+\sum x_i)}{B(a,b)} \end{aligned}$$

$$\boxed{p(x_{1:n}) = \frac{B(a+n,\, b+\sum x_i)}{B(a,b)}}$$ if $x_1, \ldots, x_n \in \{0, 1, 2, \ldots\}$, and 0 otherwise.

3. (17 points) (Exponential families, Normal distribution)
Show that the collection of $\mathcal{N}(\mu, \sigma^2)$ distributions, with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, is a two-parameter exponential family, and identify the sufficient statistics function $t(x) = (t_1(x), t_2(x))^{\mathsf{T}}$ for your parametrization.

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(- \tfrac{1}{2\sigma^2}(x-\mu)^2\big)$$

$$-\tfrac{1}{2\sigma^2}(x-\mu)^2 = -\tfrac{1}{2\sigma^2}(x^2 - 2x\mu + \mu^2) = -\tfrac{1}{2\sigma^2}x^2 + \tfrac{\mu}{\sigma^2}x - \tfrac{\mu^2}{2\sigma^2}$$

$$\begin{aligned}\mathcal{N}(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \tfrac{1}{2\sigma^2}x^2 + \tfrac{\mu}{\sigma^2}x - \tfrac{\mu^2}{2\sigma^2}\Big) \\ &= \exp\Big(- \tfrac{1}{2\sigma^2}x^2 + \tfrac{\mu}{\sigma^2}x - \tfrac{\mu^2}{2\sigma^2} - \tfrac{1}{2}\log(2\pi\sigma^2)\Big) \\ &= \exp\big(\varphi(\theta)^{\mathsf{T}}t(x) - \kappa(\theta)\big)h(x)\end{aligned}$$

where $\theta = \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}$, $\varphi(\theta) = \begin{pmatrix}-1/(2\sigma^2) \\ \mu/\sigma^2\end{pmatrix}$, $t(x) = \begin{pmatrix}x^2 \\ x\end{pmatrix}$, $\kappa(\theta) = \frac{\mu^2}{2\sigma^2} + \frac{1}{2}\log(2\pi\sigma^2)$, and $h(x) = 1$. Thus, the sufficient statistics function is $t(x) = (x^2, x)^{\mathsf{T}}$, for this choice of parametrization.

(There is more than one correct answer to this problem, since constants can be moved between $t(x)$ and $\varphi(\theta)$, as well as between $h(x)$ and $\kappa(\theta)$.)

4. (17 points) (Conjugate priors)
Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform}(0, \theta)$ given θ, that is,

$$p(x_i|\theta) = \frac{1}{\theta}\,\mathbb{1}(0 < x_i < \theta).$$

You would like to find a conjugate prior for θ. Show that the family of Pareto(α, c) distributions, with $\alpha > 0$ and $c > 0$, is a conjugate prior family.

Suppose the prior is

$$p(\theta) = \text{Pareto}(\theta|\alpha, c) = \frac{\alpha c^\alpha}{\theta^{\alpha+1}}\,\mathbb{1}(\theta > c).$$

Letting $x_* = \min\{x_1, \ldots, x_n\}$ and $x^* = \max\{x_1, \ldots, x_n\}$,

$$\begin{aligned} p(x_{1:n}|\theta) &= \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} (1/\theta)\mathbb{1}(0 < x_i < \theta) \\ &= (1/\theta^n)\mathbb{1}(0 < x_i < \theta \text{ for all } i) = (1/\theta^n)\mathbb{1}(x_* > 0,\ x^* < \theta). \end{aligned}$$

The posterior is

$$\begin{aligned} p(\theta|x_{1:n}) &\propto p(x_{1:n}|\theta)p(\theta) \\ &= (1/\theta^n)\mathbb{1}(x_* > 0,\ x^* < \theta)\frac{\alpha c^\alpha}{\theta^{\alpha+1}}\,\mathbb{1}(\theta > c) \\ &\propto \frac{1}{\theta^{\alpha+n+1}}\,\mathbb{1}(\theta > x^*)\mathbb{1}(\theta > c) \\ &= \frac{1}{\theta^{\alpha'+1}}\,\mathbb{1}(\theta > c') \\ &\propto \text{Pareto}(\theta|\alpha', c') \end{aligned}$$

where $\alpha' = \alpha + n$ and $c' = \max\{x^*, c\}$. Hence, $p(\theta|x_{1:n}) = \text{Pareto}(\theta|\alpha', c')$ and thus, the Pareto family is a conjugate prior.

5. (17 points) (Sampling methods)
Suppose $c > 1$ and

$$p(x) \propto \frac{1}{x} \mathbb{1}(1 < x < c).$$

(Note that $p(x)$ is proportional to this, not equal to this.) Assume you can generate $U \sim \text{Uniform}(0,1)$. Give an explicit formula, in terms of c and U, for generating a sample from $p(x)$. You must show your work to receive full credit.

Use the inverse c.d.f. method. First, we need to find the normalization constant. For any $b \in [1, c]$,

$$\int_1^b (1/x)dx = \log x\Big|_1^b = \log b - \log 1 = \log b.$$

Thus,

$$p(x) = \frac{1}{x \log c} \mathbb{1}(1 < x < c).$$

For any $b \in [1, c]$, the c.d.f. is therefore

$$F(b) = \int_1^b p(x)dx = \int_1^b \frac{1}{x \log c} dx = \frac{\log b}{\log c}.$$

Solving for the inverse c.d.f., for any $u \in (0, 1)$,

$$\begin{aligned} u = F(x) &= \frac{\log x}{\log c} \\ u \log c &= \log x \\ \exp(u \log c) &= x \\ c^u &= x \end{aligned}$$

Hence, **if $U \sim \text{Uniform}(0,1)$, then $c^U \sim p(x)$.**

6. (17 points) (Decision theory)
Consider a decision problem in which the state is $\theta \in \mathbb{R}$, the observation is x, you must choose an action $\hat{\theta} \in \mathbb{R}$, and the loss function is

$$\ell(\theta, \hat{\theta}) = a\theta^2 + b\theta\hat{\theta} + c\hat{\theta}^2$$

for some known $a, b, c \in \mathbb{R}$ with $c > 0$. Suppose you have computed the posterior distribution and it is $p(\theta|x) = \mathcal{N}(\theta|M, L^{-1})$ for some M and L. What is the Bayes procedure (minimizing posterior expected loss)?

(Your answer must be an explicit expression in terms of a, b, c, M, and L. You must show your work to receive full credit.)

The Bayes procedure is the decision procedure (method of choosing $\hat{\theta}$ based on x) that minimizes the posterior expected loss,

$$\begin{aligned}\rho(\hat{\theta}, x) &= \mathbb{E}\big(\ell(\boldsymbol{\theta}, \hat{\theta}) \bigm| x\big) \\ &= \mathbb{E}\big(a\boldsymbol{\theta}^2 + b\boldsymbol{\theta}\hat{\theta} + c\hat{\theta}^2 \bigm| x\big) \\ &= a\mathbb{E}(\boldsymbol{\theta}^2|x) + b\mathbb{E}(\boldsymbol{\theta}|x)\hat{\theta} + c\hat{\theta}^2.\end{aligned}$$

Since $c > 0$, this is a strictly convex quadratic function of $\hat{\theta}$, so we can set the derivative equal to zero and solve to find the minimum:

$$\begin{aligned}0 &= \frac{d}{d\hat{\theta}}\rho(\hat{\theta}, x) = b\mathbb{E}(\boldsymbol{\theta}|x) + 2c\hat{\theta} \\ \hat{\theta} &= -b\mathbb{E}(\boldsymbol{\theta}|x)/(2c).\end{aligned}$$

Since the posterior is Normal with mean M, then $\mathbb{E}(\boldsymbol{\theta}|x) = M$, hence the Bayes procedure is

$$\boxed{\hat{\theta} = \frac{-bM}{2c}.}$$