Assignment 3: Exponential families, conjugacy, and entropy

Stat 348, Spring 2024

Instructions: Submit a well-formatted LaTeX document with your answers, using the same (sub)section names as this document. Show every step—your answers should contain complete proofs.

Due: Sunday, April 14 at 11:59PM on GradeScope.

Problem 1: Exponential-gamma conjugacy

For this problem, use the following facts.

The exponential distribution with rate $\mu > 0$ has PDF

$$P(x \mid \mu) = \mathrm{Expon}(x; \mu) = \mu \exp(-\mu x), \;\; x > 0 \tag{1}$$

The gamma distribution with shape $a > 0$ and rate $b > 0$ has PDF

$$P(x \mid a, b) = \mathrm{Gam}(x; a,\, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \;\; x > 0 \tag{2}$$

In general, an exponential family distribution takes the following form:

$$P(x \mid \eta) = h(x) \exp\left(\eta^\top t(x) - a(\eta)\right) \tag{3}$$

where $h(x)$ is the base measure, η are the natural parameters, $t(x)$ are the sufficient statistics, and $a(\eta)$ is the log-normalizer.

A conjugate prior for η is also exponential family and takes the following form

$$P(\eta \mid \lambda) = h_c(\eta) \exp\left(\lambda_1^\top \eta - \lambda_2\, a_\ell(\eta) - a_c(\lambda)\right) \tag{4}$$

where $\lambda = [\lambda_1, \lambda_2]$ is the natural parameter for the conjugate prior, $t(\eta) = [\eta, -a_\ell(\eta)]$ are its sufficient statistics, and $a_c(\lambda)$ is its log-normalizer. Note that $a_\ell(\eta)$ is the log-normalizer of the likelihood $P(x \mid \eta)$, while $a_c(\lambda)$ is the log-normalizer of the prior $P(\eta \mid \lambda)$.

1.1: Exponential family forms [10 points]

Provide exponential family forms for the exponential and gamma distributions. (This means defining $h(x)$, $t(x)$, η, and $a(\eta)$ and confirming that Eq. (3) equals the PDFs above.)

1.2: Conjugacy [10 points]

Use Eq. (4), to show that if the likelihood $P(x \mid \mu)$ is exponential (Eq. (1)), then the conjugate prior for μ is a gamma distribution $P(\mu \mid a, b)$ (Eq. (2)).

1.3: Posterior [10 points]

Provide the form of the gamma posterior $P(\mu \mid x_{1:n}, a, b)$ where $x_{1:n} \equiv x_1, \ldots, x_n$, and $x_i \overset{\text{iid}}{\sim} P(x \mid \mu)$.

1.4: Prior predictive distribution [10 points]

Derive the prior predictive distribution $P(x_1 \mid a, b) = \int P(x_1 \mid \mu) P(\mu \mid a, b)\, d\mu$ for one data point x_1.

1.5: Posterior predictive distribution [10 points]

Derive the posterior predictive distribution $P(x_{n+1} \mid x_{1:n}, a, b) = \int P(x_{n+1} \mid \mu) P(\mu \mid x_{1:n}, a, b)\, d\mu$ for a single new point x_{n+1} conditional on a data set $x_{1:n}$.

Problem 2: Gamma-Poisson and entropy

For this problem, use the following facts.

The Poisson distribution with rate $\mu > 0$ has PMF

$$P(x \mid \mu) = \text{Pois}(x; \mu) = \frac{\mu^x}{x!} \exp(-\mu), \;\; x \in \mathbb{N}_0 \tag{5}$$

The negative binomial distribution with shape $r > 0$ and probability parameter $p \in (0, 1)$ has PMF

$$P(x \mid r, p) = \text{NB}(x; r, p) = \frac{\Gamma(x+r)}{x!\Gamma(r)}(1-p)^r p^x, \;\; x \in \mathbb{N}_0 \tag{6}$$

A gamma-Poisson mixture is equal to a negative binomial

$$\int_0^\infty \text{Pois}(x; \mu) \, \text{Gam}(\mu; a, \, b) \, d\mu = \text{NB}(x; a, \, \tfrac{1}{1+b}) \tag{7}$$

2.1: Posterior [5 points]

If $x \sim \text{Pois}(\mu)$ and $\mu \sim \text{Gam}(a, b)$, what is the posterior $P(\mu \mid x, a, b)$?

2.2: Exponential family forms [10 points]

Provide an exponential family form for the Poisson distribution.

In addition, provide an exponential family form for the negative binomial distribution *with known* r—this means only p is treated as a parameter, while r is treated as a known constant (e.g., like π or e).

2.3: KL divergence between two Poissons [10 points]

Use the exponential family form of the Poisson to derive the Kullback-Leibler (KL) divergence between two Poisson distributions, $\text{KL}(\text{Pois}(x; \mu_1) \,||\, \text{Pois}(x; \mu_2))$. Use the natural-log (ln) form of KL divergence: $\text{KL}(P(x) \,||\, Q(x)) = \sum_{x \in \mathscr{X}} P(x) \ln\left[\frac{P(x)}{Q(x)}\right]$.

2.4: Poisson and negative binomial entropies [25 points]

Define the following three quantities.

The entropy of a negative binomial distribution with shape a and probability parameter $\frac{1}{1+b}$:

$$H\Big(\text{NB}(a, \tfrac{1}{1+b})\Big) = -\sum_{x=0}^{\infty} \text{NB}(x; \, a, \, \tfrac{1}{1+b}) \ln\Big[\text{NB}(x; \, a, \, \tfrac{1}{1+b})\Big] \tag{8}$$

The entropy of a Poisson distribution with rate y:

$$H\Big(\text{Pois}(y)\Big) = -\sum_{x=0}^{\infty} \text{Pois}(x; y) \ln\left[\text{Pois}(x; y)\right] \tag{9}$$

The conditional entropy of a Poisson, conditioned on a gamma prior over y with shape a and rate b:

$$H\Big(\text{Pois}(y) \mid \text{Gam}(a, b)\Big) = \int_0^\infty H\Big(\text{Pois}(y)\Big) \, \text{Gam}(y; a, b) \, dy \tag{10}$$

Show that the entropy of the negative binomial is lower bounded:

$$H\Big(\text{NB}(a, \tfrac{1}{1+b})\Big) \geq H\Big(\text{Pois}(y) \mid \text{Gam}(a, b)\Big). \tag{11}$$

Problem 3: Where should we search next?

In this problem we are back to searching for the missing USS *Scorpion*. We assume the Scorpion is in one of the K search cells, and we denote its unknown location as $Z \in \{1, \ldots, K\}$. Our current beliefs about its position are encoded in the categorical distribution $P(Z = k) \equiv \pi_k$, where $\sum_{k=1}^{K} \pi_k = 1$.

We are considering which of the K search cells to send divers to next. If we send divers to cell k, we initiate a search for the sub which either succeeds or feels. Define the following binary variable

$$Y_k = \begin{cases} 1 \text{ if the search in cell } k \text{ finds the sub} \\ 0 \text{ if the search in cell } k \text{ fails to find the sub} \end{cases} \tag{12}$$

For now, assume the following:

$$P(Y_k = 1 \mid Z \neq k) = 0 \tag{13}$$

$$P(Y_k = 1 \mid Z = k) = 1 \tag{14}$$

3.1: Minimizing uncertainty [10 points]

Now we want to choose which of the possible searches $Y_1, \ldots, Y_K$ to initiate. A natural choice is the one that maximally reduces our uncertainty about Z in expectation—i.e.:

$$k^\star = \operatorname*{argmax}_k \mathrm{H}(Z) - \mathrm{H}(Z \mid Y_k) \tag{15}$$

Show that, using this selection criterion, the optimal cell to search next is equal to

$$k^\star = \operatorname*{argmax}_k \pi_k \tag{16}$$

3.2: Incorporating SEPs [15 points]

Now assume the following:

$$P(Y_k = 1 \mid Z \neq k) = 0 \tag{17}$$

$$P(Y_k = 1 \mid Z = k) = q_k \tag{18}$$

where $q_k \in [0, 1]$ is the search effectiveness probability (SEP) of cell k. Taking q_k into account, which search $k^\star$ would minimize our uncertainty about Z? In other words, solve again for:

$$k^\star = \operatorname*{argmax}_k \mathrm{H}(Z) - \mathrm{H}(Z \mid Y_k) \tag{19}$$

Your answer should be in the form $k^\star = \operatorname*{argmax}_k f(\pi_k, q_k)$ where $f(\cdots)$ is a simple function of π_k and q_k. You may use the *binary entropy function* $\mathrm{H}_2(p) \triangleq p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$ in your answer, but otherwise provide a form for $f(\cdots)$ that is as simplified as possible.

3.3: Example involving SEPs [10 points]

Now consider $K = 4$, and the probabilities π_k and SEPs q_k equal to the following:

$$\pi = [\tfrac{3}{4}, \tfrac{1}{8}, \tfrac{1}{16}, \tfrac{1}{16}] \tag{20}$$

$$q = [\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{3}{4}, \tfrac{1}{2}] \tag{21}$$

Using these values, and the form you derived above for $k^\star = f(\pi_k, q_k)$, provide $k^\star$ (you may use a calculator). Is this the answer you expected? Provide some reflection on the answer, and what it tells you about the relationship between SEPs and the optimal search, in terms of uncertainty reduction.