

Assignment 3: Exponential families, conjugacy, and entropy

Stat 348, Spring 2023

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

Due: Monday, April 17 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

$$P(x | \mu) = \text{Expon}(x; \mu) = \mu \exp(-\mu x), \quad x > 0 \quad (1)$$

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

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

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

$$P(x | \eta) = h(x) \exp(\eta^\top t(x) - a(\eta)) \quad (3)$$

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

$$P(\eta | \lambda) = h_c(\eta) \exp(\lambda_1^\top \eta - \lambda_0 a_\ell(\eta) - a_c(\lambda)) \quad (4)$$

where $\lambda = [\lambda_1, \lambda_0]$ are the natural parameters 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 $P(x | \eta)$, while $a_c(\lambda)$ is the log-normalizer of the prior $P(\eta | \lambda)$.

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 $P(x | \mu)$ is exponential (Eq. (1)), then the conjugate prior for μ is a gamma distribution $P(\mu | a, b)$ (Eq. (2)).

1.3: Posterior [10 points]

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

1.4: Prior predictive distribution [10 points]

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

1.5: Posterior predictive distribution [10 points]

Derive the posterior predictive distribution $P(x_{n+1} | x_{1:n}, a, b) = \int P(x_{n+1} | \mu) P(\mu | x_{1:n}, a, b) d\mu$ 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

$$P(x | \mu) = \text{Pois}(x; \mu) = \frac{\mu^x}{x!} \exp(-\mu), \quad x \in \{1, 2, \dots\} \quad (5)$$

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

$$P(x | r, p) = \text{NB}(x; r, p) = \frac{\Gamma(x+r)}{x! \Gamma(r)} (1-p)^r p^x, \quad x \in \{1, 2, \dots\} \quad (6)$$

A gamma-Poisson mixture is equal to a negative binomial

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

2.1: Posterior [5 points]

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

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: $\text{KL}(P(x) || Q(x)) = \sum_{x \in \mathcal{X}} P(x) \ln \left[\frac{P(x)}{Q(x)} \right]$.

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 $\frac{1}{1+b}$:

$$H\left(\text{NB}(a, \frac{1}{1+b})\right) = - \sum_{x=0}^{\infty} \text{NB}(x; a, \frac{1}{1+b}) \ln [\text{NB}(x; a, \frac{1}{1+b})] \quad (8)$$

The entropy of a Poisson distribution with rate y :

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

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

$$H\left(\text{Pois}(y) | \text{Gam}(a, b)\right) = \int_0^\infty H\left(\text{Pois}(y)\right) \text{Gam}(y; a, b) dy \quad (10)$$

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

$$H\left(\text{NB}(a, \frac{1}{1+b})\right) \geq H\left(\text{Pois}(y) | \text{Gam}(a, b)\right). \quad (11)$$