Homework 1

Aukkawut Ammartayakun
CS 539 Machine Learning

Spring 2023

Problem 1

Consider two variables X and Y with joint distribution $p(x, y)$. Prove the following two results

1. $\mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}_X[X|Y]\right]$

   Solution. Consider the RHS,

$$\begin{aligned}
\mathbb{E}_Y\left[\mathbb{E}_X[X|Y]\right] &= \int_{\Omega_Y} f_Y(y)\, dy \\
&= \int_{\Omega_Y}\int_{\Omega_X} x f_{X|Y}(x|y) f_Y(y)\, dx\, dy \\
&= \int_{\Omega_Y}\int_{\Omega_X} x p(x,y)\, dx\, dy \\
&= \int_{\Omega_X}\int_{\Omega_Y} x p(x,y)\, dy\, dx \\
&= \int_{\Omega_X} x f_X(x)\, dx \\
&= \mathbb{E}[X]
\end{aligned}$$

2. $\mathrm{Var}[X] = \mathbb{E}_Y\left[\mathrm{Var}_X[X|Y]\right] + \mathrm{Var}_Y\left[\mathbb{E}_X[X|Y]\right]$

   Solution. Consider the LHS,

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}^2[X] \\
&= \mathbb{E}\left[\mathrm{Var}[X|Y] + \mathbb{E}^2[X|Y]\right] - \mathbb{E}^2[\mathbb{E}[X|Y]] \\
&= \mathbb{E}\left[\mathrm{Var}[X|Y]\right] + \mathbb{E}[\mathbb{E}^2[X|Y]] - \mathbb{E}^2[\mathbb{E}[X|Y]] \\
&= \mathbb{E}_Y\left[\mathrm{Var}_X[X|Y]\right] + \mathrm{Var}_Y\left[\mathbb{E}_X[X|Y]\right]
\end{aligned}$$

Problem 2

Let X and Z be two independent random vectors, so that $p(X, Z) = p(X)p(Z)$. Show that the mean of their sum $Y = X + Z$ is given by the sum of the means of each of the variables separately. Similarly, show that the covariance matrix of Y is given by the sum of the covariance matrices of X and Z.

Solution. Consider the mean of joint distribution $Y = X + Z$,

$$
\begin{aligned}
\mathbb{E}[Y] &= \int_{\Omega_X} \int_{\Omega_Z} y f_Y(y)\, dy\, dx \\
&= \int_{\Omega_X} \int_{\Omega_Z} (x+z) f_X(x) f_Z(z)\, dx\, dz \\
&= \int_{\Omega_X} \int_{\Omega_Z} x f_X(x) f_Z(z)\, dx\, dz + \int_{\Omega_X} \int_{\Omega_Z} z f_X(x) f_Z(z)\, dx\, dz \\
&= \int_{\Omega_X} x f_X(x)\, dx + \int_{\Omega_Z} z f_Z(z)\, dz \\
&= \mathbb{E}[X] + \mathbb{E}[Z]
\end{aligned}
$$

And also consider the covariance matrix of joint distribution $Y = X + Z$,

$$
\begin{aligned}
\text{cov}[Y] &= \mathbb{E}\left[(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^T\right] \\
&= \mathbb{E}\left[(X + Z - \mathbb{E}[X] - \mathbb{E}[Z])(X + Z - \mathbb{E}[X] - \mathbb{E}[Z])^T\right] \\
&= \mathbb{E}\left[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\right] + \mathbb{E}\left[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^T\right] \\
&= \text{cov}[X] + \text{cov}[Z]
\end{aligned}
$$

Problem 3

Consider a D-dimensional Gaussian random variable X with distribution $\mathcal{N}(x|\mu, \Sigma)$ in which the covariance Σ is known and for which we wish to infer the mean μ from a set of observations $\mathcal{X} = \{X_1, \ldots, X_n\}$. Given a prior distribution $p(\mu) = \mathcal{N}(\mu|\mu_0, \Sigma_0)$, find the corresponding posterior distribution $p(\mu|\mathcal{X})$

Problem 4

Consider a univariate Gaussian distribution $\mathcal{N}(x|\mu, \tau^{-1})$ having conjugate Gaussian-gamma prior given by

$$p(\mu, \lambda) = \mathcal{N}\left(\mu|\mu_0, (\beta\lambda)^{-1}\right) \text{Gamma}(\lambda|a, b)$$

and a dataset $\mathcal{X} = \{X_1, \ldots, X_n\}$ of i.i.d. observations. Show that the posterior distribution is also a Gaussian-gamma distribution of the same functional form as the prior, and write down expressions for the parameters of this posterior distribution.

Problem 5

Show that if two variables X and Y are independent, then their covariance is zero.

Lemma 5.1. *Let X and Y be random variable where $X \perp\!\!\!\perp Y$, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$*

Proof. Assume that $X \sim \mu(x)$ and $Y \sim \nu(y)$ where $\mu : \Omega_X \to \mathcal{B}_X$ and $\nu : \Omega_Y \to \mathcal{B}_Y$ where $\mathcal{B}$ is a Borel field, the expectation operator is defined as

$$\mathbb{E}[X] = \int_{\Omega_X} x \, d\mu(x)$$

Using the Fubini-Tonelli theorem, we can show that

$$\begin{aligned}\mathbb{E}[XY] &= \int_{\Omega_X}\int_{\Omega_Y} xy \, d\pi(x,y) \\ &= \int_{\Omega_X} x \, d\mu(x) \int_{\Omega_X} y \, d\nu(y) \\ &= \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

□

Solution. We will show that if $X \perp\!\!\!\perp Y$, then $\mathrm{cov}(X,Y) = 0$. Recall that the definition of covariance is

$$\mathrm{cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] \tag{1}$$

From this, we can simplify the covariance as

$$\begin{aligned}\mathrm{cov}(X,Y) &= \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] \\ &= \mathbb{E}\left[XY - Y\mathbb{E}[X]Y - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y])\right] \\ &= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] - [X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]) \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

Since, $X \perp\!\!\!\perp Y$. That implies $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ (from lemma 5.1). In other words, $\mathrm{cov}(X,Y) = 0$.

Problem 6

Evaluate the Kullback-Leibler divergence between two Gaussians $p(x) = \mathcal{N}(x|\mu,\sigma^2)$ and $q(x) = \mathcal{N}(x|m,s^2)$

Solution. Recall the definition of KL divergence between two measures $p(x)$ and $q(x)$ given that $p, q : \mathbb{R} \to \mathcal{B}$ where $\mathcal{B}$ is a Borel field.

$$D_{KL}(p(x)||q(x)) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx \tag{2}$$

In this case,

$$\begin{aligned}D_{KL}(p(x)||q(x)) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \ln\left(\frac{\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi}s}\exp\left(-\frac{(x-m)^2}{2s^2}\right)}\right) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\left(\ln\left(\frac{s}{\sigma}\right) + \left[\frac{(x-m)^2}{2s^2} - \frac{(x-\mu)^2}{2\sigma^2}\right]\right) dx \\ &= \ln\left(\frac{s}{\sigma}\right) + \frac{1}{\sqrt{2\pi}\sigma}\left[\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\frac{(x-m)^2}{2s^2}\,dx - \int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\frac{(x-\mu)^2}{2\sigma^2}\,dx\right] \\ &= \ln\left(\frac{s}{\sigma}\right) - \frac{1}{2} + \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\frac{(x-m)^2}{2s^2}\,dx \\ &= \ln\left(\frac{s}{\sigma}\right) - \frac{1}{2} + \frac{(\mu-m)^2+\sigma^2}{s^2}\end{aligned}$$

Problem 7

For Inverse Gamma distribution

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right)$$

1. Show the mean of the distribution is

$$\mathbb{E}[X] = \frac{\beta}{\alpha - 1}$$

Solution. We want to show that $\mathbb{E}[X] = \frac{\beta}{\alpha-1}$. Using the definition of expectation and the fact that

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x)\, dx \tag{3}$$

Let $u = \frac{\beta}{x}$

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\beta^\alpha}{\Gamma(\alpha)} \beta^{-\alpha+1} \int_0^\infty u^{\alpha-2} \exp(-u) du \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \beta^{-\alpha+1} \Gamma(\alpha - 1) \\
&= \frac{\beta}{\alpha - 1}
\end{aligned}$$

2. Find maximum likelihood estimate of α and β given N independent sample of x_i

Solution. The maximum likelihood can be evaluated from its likelihood function. Namely,

$$\mathcal{L}(\alpha, \beta | X) = \prod_{i=1}^{N} f(x_i | a, b) \tag{4}$$

In this case, the likelihood function is

$$\mathcal{L}(\alpha, \beta | X) = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \exp\left(-\sum_{i=1}^{N} \frac{\beta}{x_i}\right) \prod_{k=1}^{N} x_k^{-(\alpha+1)}$$

Minimize the log-likelihood with respect to α while fixing β

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \ln\left(\mathcal{L}(\alpha, \beta | X)\right) &= \frac{\partial}{\partial \alpha} \ln\left[\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^N \exp\left(-\sum_{i=1}^{N} \frac{\beta}{x_i}\right) \prod_{k=1}^{N} x_k^{-(\alpha+1)}\right] \\
&= \frac{\partial}{\partial \alpha} \left[N \ln\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right) - \sum_{i=1}^{N} \frac{\beta}{x_i} + \sum_{i=1}^{N} x_k^{-(\alpha+1)}\right] \\
&= \frac{\partial}{\partial \alpha} \left[N\alpha \ln(\beta) - N \ln(\Gamma(\alpha)) + \sum_{k=1}^{N} x_k^{-(\alpha+1)}\right] \\
&= N \ln(\beta) - N\psi(\alpha) + \sum_{k=1}^{N} (-x_k^{-(\alpha+1)} \ln(x_k))
\end{aligned}$$

and with respect to β while fixing α

$$\frac{\partial}{\partial \beta} \left[N\alpha \ln(\beta) - N \ln(\Gamma(\alpha)) - \sum_{i=1}^{N} \frac{\beta}{x_i}\right] = \frac{N\alpha}{\beta} - \sum_{i=1}^{N} \frac{1}{x_i}$$

Now, we can set the partial derivative to zero and solve for α and β.

$$N\ln(\beta) - N\psi(\alpha) + \sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) = 0 \tag{5}$$

$$\frac{N\alpha}{\beta} - \sum_{i=1}^{N}\frac{1}{x_i} = 0 \tag{6}$$

From the equation 6, we can see that $\beta = \frac{N\alpha}{\sum_{i=1}^{N}\frac{1}{x_i}}$. Substituting this into equation 5, we can solve for α. That is,

$$\begin{aligned}
N\ln\left(\frac{N\alpha}{\sum_{i=1}^{N}\frac{1}{x_i}}\right) - N\psi(\alpha) + \sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) &= 0\\
N\ln\left(\frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}\right) - N\psi(\alpha) + \sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) &= 0\\
\sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) &= N\psi(\alpha) - N\ln\left(\frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}\right)\\
\sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) &= N\psi(\alpha) - N\ln(N) + N\ln\left(\sum_{i=1}^{N}\frac{1}{x_i}\right)\\
\sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) &= N\psi(\alpha) - N\ln(N) + N\ln\left(\frac{1}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{x_i}}\right)\\
\sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) &= N\psi(\alpha) - N\ln(N) + N\ln\left(\frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}\right)\\
\sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) &= N\psi(\alpha) + N\ln\left(\sum_{i=1}^{N}\frac{1}{x_i}\right)\\
\psi(\alpha) &= \frac{1}{n}\sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) - \ln(\sum_{i=1}^{N}\frac{1}{x_i})\\
\alpha &= \psi^{-1}\left(\frac{1}{n}\sum_{k=1}^{N}(-x_k^{-(\alpha+1)}\ln(x_k)) - \ln(\sum_{i=1}^{N}\frac{1}{x_i})\right)
\end{aligned}$$

where ψ^{-1} is the inverse of the digamma function.

3. Discuss if $f(x)$ follows an exponential family distribution or not.

Solution. To show that $f(x|\theta)$ is the exponential family, we need to show that

$$f(x|\theta) = h(x)c(\theta)\exp(T(x)\tau(\theta))$$

From the definition of $f(x)$, we can expand it as

$$\begin{aligned}
f(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\exp\left(-\frac{\beta}{x}\right)\\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\exp\left(\ln(x^{-\alpha-1})\right)\exp\left(-\frac{\beta}{x}\right)\\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\exp\left((-\alpha-1)\ln(x) - \frac{\beta}{x}\right)
\end{aligned}$$

In this case, $T(x) = \begin{bmatrix}\ln(x) & \frac{1}{x}\end{bmatrix}$ and $\tau(\theta) = \begin{bmatrix}-\alpha - 1 & \beta\end{bmatrix}$, $c(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}$ and $h(x) = 1$. Thus, $f(x)$ is in the exponential family.

4. Discuss if X is an inverse gamma distribution with $\alpha = 1$ and $\beta = c$, then $Y = 1/X$ is an exponential distribution with rate c

   Solution. Using the CDF method, one can come to conclude that $P(X \geq 1/y) = 1 - F(\frac{1}{y})$ for $Y = y$ and $Y = 1/X$. Thus, evaluate its derivative and we can get

$$g(y) = \frac{1}{y^2} f\left(\frac{1}{y}\right) = \frac{1}{y^2}\frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha+1} \exp\left(-\beta y\right)$$

   Let $\alpha = 1, \beta = c$ we will get exponential PDF.

$$g(y) = c\exp(-cy)$$

5. For distribution Y, and with a prior on c define the posterior of c. Pick a proper prior for c first.

   Solution. Using gamma prior, one can get the posterior of c from conjugate prior,

$$\begin{aligned}
P(c|Y) &\propto P(Y|c)P(c) \\
&= c^{n+\alpha-1} \exp\left(-c\left(\left[\sum_{i=1}^{n} y\right] - \beta\right)\right) \\
P(c|Y) &\sim \text{Gamma}\left(n+\alpha, \left[\sum_{i=1}^{n} y\right] - \beta\right)
\end{aligned}$$

   That is, the posterior of c would be a gamma distribution with parameters $n + \alpha$ and $[\sum_{i=1}^{n} y] - \beta$.