Homework 7 Solutions

Chapter 11

1) Stein’s lemma.
   a) \( f_1 = N(0, \sigma_1^2), \ f_2 = N(0, \sigma_2^2), \)
   \[
   D(f_1 \| f_2) = \int_{-\infty}^{\infty} f_1(x) \left[ \frac{1}{2} \ln \frac{\sigma_2^2}{\sigma_1^2} - \left( \frac{x^2}{2\sigma_1^2} - \frac{x^2}{2\sigma_2^2} \right) \right] \, dx
   \]
   \[
   = \frac{1}{2} \left[ \ln \frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_2^2}{\sigma_1^2} - 1 \right].
   \]

   b) \( f_1 = \lambda_1 e^{-\lambda_1 x}, \ f_2 = \lambda_2 e^{-\lambda_2 x}. \)
   \[
   D(f_1 \| f_2) = \int_{0}^{\infty} f_1(x) \left[ \ln \frac{\lambda_1}{\lambda_2} - \lambda_1 x + \lambda_2 x \right] \, dx
   \]
   \[
   = \ln \frac{\lambda_1}{\lambda_2} + \frac{\lambda_2}{\lambda_1} - 1.
   \]

   c) \( f_1 = U[0, 1], \ f_2 = U[a, a+1], \)
   \[
   D(f_1 \| f_2) = \int_{0}^{1} f_1 \ln \frac{f_1}{f_2}
   \]
   \[
   = \int_{0}^{a} f_1 \ln \infty + \int_{a}^{1} f_1 \ln 1
   \]
   \[
   = \infty.
   \]

   In this case, the Kullback-Leibler distance of \( \infty \) implies that in a hypothesis test, the two distributions will be distinguished with probability 1 for large samples.

   d) \( f_1 = \text{Bern} \left( \frac{1}{2} \right) \) and \( f_2 = \text{Bern}(1), \)
   \[
   D(f_1 \| f_2) = \frac{1}{2} \ln \frac{1}{1} + \frac{1}{2} \ln \frac{1}{0} = \infty.
   \]

   The implication is the same as in part (c).

2) A relation between \( D(P \| Q) \) and Chi-square.
   There are many ways to expand \( D(P \| Q) \) in a Taylor series, but when we are expanding about \( P = Q, \) we must get a series in \( P - Q, \) whose coefficients depend on \( Q \) only. It is easy to get misled into forming another series expansion, so we will provide two alternative proofs of this result.
   
   - Expanding the log.
Writing \( P_Q = 1 + \frac{P - Q}{Q} = 1 + \frac{\Delta}{Q} \), and \( P = Q + \Delta \), we get

\[
D(P||Q) = \int P \ln \frac{P}{Q} \tag{1175}
\]

\[
= \int (Q + \Delta) \ln \left( 1 + \frac{\Delta}{Q} \right) \tag{1176}
\]

\[
= \int (Q + \Delta) \left( \frac{\Delta}{Q} - \frac{\Delta^2}{2Q^2} + \ldots \right) \tag{1177}
\]

\[
= \int \Delta + \frac{\Delta^2}{Q} - \frac{\Delta^2}{2Q} + \ldots \tag{1178}
\]

The integral of the first term \( \int \Delta = \int P - \int Q = 0 \), and hence the first non-zero term in the expansion is

\[
\frac{\Delta^2}{2Q} = \frac{\chi^2}{2}, \tag{1179}
\]

which shows that locally around \( Q \), \( D(P||Q) \) behaves quadratically like \( \chi^2 \).

- By differentiation.
  If we construct the Taylor series expansion for \( f \), we can write

\[
f(x) = f(c) + f'(c)(x - c) + f''(c)\frac{(x - c)^2}{2} + \ldots \tag{1180}
\]

Doing the same expansion for \( D(P||Q) \) around the point \( Q \), we get

\[
D(P||Q)_{P=Q} = 0, \tag{1181}
\]

\[
D'(P||Q)_{P=Q} = (\ln \frac{P}{Q} + 1)_{P=Q} = 1, \tag{1182}
\]

and

\[
D''(P||Q)_{P=Q} = \left( \frac{1}{P} \right)_{P=Q} = \frac{1}{Q}, \tag{1183}
\]

Hence the Taylor series is

\[
D(P||Q) = 0 + \int 1(P - Q) + \int \frac{1}{Q} \frac{(P - Q)^2}{2} + \ldots \tag{1184}
\]

\[
= \frac{1}{2} \chi^2 + \ldots \tag{1185}
\]

and we get \( \chi^2 \) as the first non-zero term in the expansion.
3) Error exponent for universal codes.

a) We have to minimize $D(p||q)$ subject to the constraint that $H(p) \geq R$. Rewriting this problem using Lagrange multipliers, we get

$$J(p) = \sum_p p \log \frac{p}{q} + \lambda \sum p \log p + \nu \sum p.$$  \hspace{1cm} (1186)

Differentiating with respect to $p(x)$ and setting the derivative to 0, we obtain

$$\log \frac{p}{q} + 1 + \lambda \log p + \lambda + \nu = 0,$$  \hspace{1cm} (1187)

which implies that

$$p^*(x) = \frac{q^\mu(x)}{\sum_a q^\mu(a)},$$  \hspace{1cm} (1188)

where $\mu = \frac{\lambda}{1 - \lambda}$ is chosen to satisfy the constraint $H(p^*) = R$. We have to first check that the constraint is active, i.e., that we really need equality in the constraint. For this we set $\lambda = 0$ or $\mu = 1$, and we get $p^* = q$. Hence if $q$ is such that $H(q) \geq R$, then the maximizing $p^*$ is $q$. On the other hand, if $H(q) < R$, then $\mu \neq 0$, and the constraint must be satisfied with equality.

Geometrically it is clear that there will be two solutions for $\lambda$ of the form (1188) which have $H(p^*) = R$, corresponding to the minimum and maximum distance to $q$ on the manifold $H(p) = R$. It is easy to see that for $0 \leq \mu \leq 1$, $p^*_\mu(x)$ lies on the geodesic from $q$ to the uniform distribution. Hence, the minimum will lie in this region of $\mu$. The maximum will correspond to negative $\mu$, which lies on the other side of the uniform distribution as in the figure.

b) For a universal code with rate $R$, any source can be transmitted by the code if $H(p) < R$. In the binary case, this corresponds to $p \in [0, h^{-1}(R))$ or $p \in (1 - h^{-1}(R), 1]$, where $h$ is the binary entropy function.
7) Fisher information and relative entropy. Let \( t = \theta' - \theta \). Then
\[
\frac{1}{(\theta - \theta')^2} D(p_\theta||p_{\theta'}) = \frac{1}{t^2} D(p_\theta||p_{\theta+t}) = \frac{1}{t^2 \ln 2} \sum_x p_\theta(x) \ln \frac{p_\theta(x)}{p_{\theta+t}(x)}. \tag{1245}
\]
Let
\[
f(t) = \frac{p_\theta(x) \ln \frac{p_\theta(x)}{p_{\theta+t}(x)}}{p_{\theta+t}(x)}. \tag{1246}
\]
We will suppress the dependence on \( x \) and expand \( f(t) \) in a Taylor series in \( t \). Thus
\[
f'(t) = -\frac{p_\theta}{p_{\theta+t}} \frac{dp_{\theta+t}}{dt}, \tag{1247}
\]
and
\[
f''(t) = \frac{p_\theta}{p_{\theta+t}} \left( \frac{dp_{\theta+t}}{dt} \right)^2 + \frac{p_\theta}{p_{\theta+t}} \frac{d^2 p_{\theta+t}}{dt^2}. \tag{1248}
\]
Thus expanding in the Taylor series around \( t = 0 \), we obtain
\[
f(t) = f(0) + f'(0)t + f''(0)\frac{t^2}{2} + O(t^3), \tag{1249}
\]
where \( f(0) = 0 \),
\[
f'(0) = -\frac{p_\theta}{p_\theta} \left. \frac{dp_\theta}{dt} \right|_{t=0} = \frac{dp_\theta}{d\theta} \tag{1250}
\]
and
\[
f''(0) = \frac{1}{p_\theta} \left( \frac{dp_\theta}{d\theta} \right)^2 + \frac{d^2 p_\theta}{d\theta^2} \tag{1251}
\]
Now \( \sum_x p_\theta(x) = 1 \), and therefore
\[
\sum_x \frac{dp_\theta(x)}{d\theta} = \frac{d}{dt} 1 = 0, \tag{1252}
\]
and
\[
\sum_x \frac{d^2 p_\theta(x)}{d\theta^2} = \frac{d}{dt} 0 = 0. \tag{1253}
\]
Therefore the sum of the terms of (1250) sum to 0 and the sum of the second terms in (1251) is 0. Thus substituting the Taylor expansions in the sum, we obtain
\[
\frac{1}{(\theta - \theta')^2} D(p_\theta||p_{\theta'}) = \frac{1}{t^2 \ln 2} \sum_x p_\theta(x) \ln \frac{p_\theta(x)}{p_{\theta+t}(x)} \tag{1254}
\]
\[
= \frac{1}{t^2 \ln 2} \left( 0 + \sum_x \frac{dp_\theta(x)}{d\theta} t + \sum_x \left( \frac{1}{p_\theta} \left( \frac{dp_\theta}{d\theta} \right)^2 + \frac{d^2 p_\theta}{d\theta^2} \right) \frac{t^2}{2} + O(t^3) \right) \tag{1255}
\]
\[
= \frac{1}{2 \ln 2} \sum_x \frac{1}{p_\theta(x)} \left( \frac{dp_\theta(x)}{d\theta} \right)^2 + O(t) \tag{1256}
\]
\[
= \frac{1}{\ln 4} J(\theta) + O(t) \tag{1257}
\]
and therefore
\[
\lim_{t \to 0} \frac{1}{(\theta - \theta')^2} D(p_\theta||p_{\theta'}) = \frac{1}{\ln 4} J(\theta). \tag{1258}
\]
13) Sanov’s theorem

- Since \( nX/n \) has a binomial distribution, we have

\[
\Pr(nX/n = i) = \binom{n}{i} q^i (1 - q)^{n-i}
\]

and therefore

\[
\Pr\{X_1, X_2, \ldots, X_n : pX \geq p\} \leq \sum_{i = [np]}^{n} \binom{n}{i} q^i (1 - q)^{n-i}
\]

This ratio is less than 1 if \( \frac{n - i}{i+1} < \frac{1 - q}{q} \) i.e., if \( i > nq - (1 - q) \). Thus the maximum of the terms occurs when \( i = \lfloor np \rfloor \).

- From Example 11.1.3,

\[
\binom{n}{\lfloor np \rfloor} \geq 2^n H(p)
\]

and hence the largest term in the sum is

\[
\binom{n}{\lfloor np \rfloor} q^{\lfloor np \rfloor} (1 - q)^{n - \lfloor np \rfloor} 
\geq 2^n (-p \log p - (1-p) \log (1-p)) + np \log q + n(1-p) \log (1-q) - 2^{-nD(p||q)}
\]

- From the above results, it follows that

\[
\Pr\{X_1, X_2, \ldots, X_n : pX \geq p\} \leq \sum_{i = [np]}^{n} \binom{n}{i} q^i (1 - q)^{n-i}
\]

\[
\leq \binom{n - \lfloor np \rfloor}{\lfloor np \rfloor} \binom{n}{\lfloor np \rfloor} q^i (1 - q)^{n-i}
\]

\[
\leq (n(1 - p) + 1) 2^{-nD(p||q)}
\]

where the second inequality follows from the fact that the sum is less than the largest term times the number of terms. Taking the logarithm and dividing by \( n \) and taking the limit as \( n \to \infty \), we obtain

\[
\lim_{n \to \infty} \frac{1}{n} \log \Pr\{X_1, X_2, \ldots, X_n : pX \geq p\} \leq -D(p||q)
\]

Similarly, using the fact the sum is larger than the largest term, we obtain

\[
\Pr\{X_1, X_2, \ldots, X_n : pX \geq p\} \geq \sum_{i = [np]}^{n} \binom{n}{i} q^i (1 - q)^{n-i}
\]

\[
\geq \binom{n}{\lfloor np \rfloor} q^i (1 - q)^{n-i}
\]

\[
\geq 2^{-nD(p||q)}
\]

and

\[
\lim_{n \to \infty} \frac{1}{n} \log \Pr\{X_1, X_2, \ldots, X_n : pX \geq p\} \geq -D(p||q)
\]

Combining these two results, we obtain the special case of Sanov’s theorem

\[
\lim_{n \to \infty} \frac{1}{n} \log \Pr\{X_1, X_2, \ldots, X_n : pX \geq p\} = -D(p||q)
\]