Problem 1

American Football

a) Note that $\mathcal{D}(\tilde{Q} \parallel Q_{\text{pass}})$ will be infinite if there is any symbol $z \in \mathbb{N}_0$ with $Q_{\text{pass}}(z) = 0$, but $\tilde{Q}(z) > 0$. Hence, to minimize the relative entropy, we only consider $z \in \{0, 10\}$. Let $\tilde{Q}(10) = 1 - \tilde{Q}(0) = p$. Then,

$$E_\tilde{Q}[Z] = 10p + 0 \cdot (1 - p) = 10p \geq 4 \implies p \geq \frac{2}{5}.$$  

Moreover,

$$\mathcal{D}(\tilde{Q}^* \parallel Q) = p \log \frac{p}{5} + (1 - p) \log \frac{1 - p}{4}$$

$$= - H_b(p) + p \log 4 + \log \frac{5}{4};$$

$$\frac{\partial}{\partial p} \mathcal{D}(\tilde{Q}^* \parallel Q) = \log \frac{p}{1 - p} + \log 4 = 0 \implies p = \frac{1}{5} < \frac{2}{5}.$$  

Hence, the minimum value of the relative entropy under the condition that $E[Z] \geq 4$ is achieved for $p = \frac{2}{5}$:

$$\tilde{Q}^*(z) = \begin{cases} \frac{3}{5} & \text{for } z = 0, \\ \frac{2}{5} & \text{for } z = 10 \end{cases}$$

and

$$\mathcal{D}(\tilde{Q}^* \parallel Q) = \frac{3}{5} \log \frac{3}{5} + \frac{2}{5} \log \frac{2}{5} = \frac{1}{5} \log \frac{27}{16} \approx 0.105 \text{ nats.}.$$  

We could also have used two Lagrange multipliers:

$$L(\tilde{Q}) = \sum_{z \in \{0, 10\}} \tilde{Q}(z) \log \frac{\tilde{Q}(z)}{Q_{\text{pass}}(z)} - \lambda_1 \left( \sum_{z \in \{0, 10\}} z \tilde{Q}(z) - 4 \right)$$

$$- \lambda_2 \left( \sum_{z \in \{0, 10\}} \tilde{Q}(z) - 1 \right);$$

$$\frac{\partial L(\tilde{Q}(b))}{\partial \tilde{Q}(b)} = \log \frac{\tilde{Q}(b)}{Q_{\text{pass}}(b)} + 1 - \lambda_1 b - \lambda_2 = 0$$

$$\implies \tilde{Q}^*(z) = Q_{\text{pass}}(z) e^{\lambda_2 - 1 + \lambda_1 z}.$$
The parameters $\lambda_1$ and $\lambda_2$ can be computed as follows:

\[
\sum_{z \in \{0,10\}} \tilde{\mathcal{Q}}^*(z) = e^{\lambda_2 - 1} \left( \frac{4}{5} + \frac{1}{5} e^{10\lambda_1} \right) = 1
\]

\[
\Rightarrow e^{\lambda_2 - 1} = \frac{5}{4 + e^{10\lambda_1}};
\]

\[
\sum_{z \in \{0,10\}} z \tilde{\mathcal{Q}}^*(z) = \frac{5}{4 + e^{10\lambda_1}} \frac{10}{5} e^{10\lambda_1} = 4
\]

\[
\Rightarrow 10 e^{10\lambda_1} = 16 + 4 e^{10\lambda_1}
\]

\[
\Rightarrow e^{10\lambda_1} = \frac{8}{3}
\]

\[
\Rightarrow \lambda_1 = \frac{1}{10} \log \frac{8}{3},
\]

which yields again (1).

b) Again note that $\mathcal{D}(\mathcal{Q} | Q_{\text{run}})$ will be infinity if there is any symbol $z \in \mathbb{N}_0$ where $Q_{\text{run}}(z) = 0$, but $\mathcal{Q}(z) > 0$. Hence, only consider $z \in \{1,2,3,4\}$. But any distribution on $\{1,2,3,4\}$ can only have an expectation larger or equal to 4 if the symbols 1, 2, and 3 have zero probability! Hence,

\[
\tilde{\mathcal{Q}}^*(z) = \begin{cases} 1 \text{ for } z = 4, \\ 0 \text{ otherwise.} \end{cases}
\]

The relative entropy is then

\[
\mathcal{D}(\tilde{\mathcal{Q}}^* || Q) = \log \frac{1}{4} = \log 4 \approx 1.39 \text{ nats.}
\]

c) Sanov’s theorem tells us that the probability of this event is about

\[
\exp \left( -n \inf_{\tilde{\mathcal{Q}} \in \mathcal{P}(\mathbb{N}_0)}: \mathbb{E}_{\tilde{\mathcal{Q}}}[Z] \geq 4 \mathcal{D}(\tilde{\mathcal{Q}} || \mathcal{Q}_{\text{pass}}) \right) = \left( \frac{16}{27} \right)^{n/5} \approx 0.90^n.
\]

d) Similarly, Sanov’s theorem tells us that the probability of this event is about

\[
\exp \left( -n \inf_{\tilde{\mathcal{Q}} \in \mathcal{P}(\mathbb{N}_0)}: \mathbb{E}_{\tilde{\mathcal{Q}}}[Z] \geq 4 \mathcal{D}(\tilde{\mathcal{Q}} || \mathcal{Q}_{\text{run}}) \right) = \left( \frac{1}{4} \right)^n = 0.25^n.
\]

So clearly the strategy of passing only has a much higher probability of success.

Problem 2

Sanov-Type Theorem for the Size of Type Classes

Note that any sequence $x \in \mathcal{X}^n$ satisfying $\frac{1}{n} \sum_{k=1}^{n} g(x_k) \geq \alpha$ must have a type $P_x$ that satisfies $\sum_{a \in \mathcal{X}} P_x(a) g(a) \geq \alpha$. Hence the size of

\[
\mathcal{A} = \left\{ x \in \mathcal{X}^n : \frac{1}{n} \sum_{k=1}^{n} g(x_k) \geq \alpha \right\}
\]

is identical to $|\mathcal{T}^n(\mathcal{F})|$ where

\[
\mathcal{F} = \left\{ \mathcal{Q} \in \mathcal{P}(\mathcal{X}) : \sum_{a \in \mathcal{X}} Q(a) g(a) \geq \alpha \right\}.
\]

© Stefan M. Moser, 2015
So let’s prove the theorem. We note that $\mathcal{P}_n(\mathcal{X})$ is finite, i.e., we can find a $P^* \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})$ such that

$$H(P^*) = \max_{P \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} H(P).$$

We start with the upper bound:

$$|T^n(\mathcal{F})| = |T^n(\mathcal{F} \cap \mathcal{P}_n(\mathcal{X}))|$$

$$= \sum_{P \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} |T^n(P)|$$

$$\leq \sum_{P \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} \max_{\tilde{P} \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} |T^n(\tilde{P})|$$

$$= \max_{\tilde{P} \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} |T^n(\tilde{P})| \cdot |\mathcal{F} \cap \mathcal{P}_n(\mathcal{X})|$$

$$\leq \max_{\tilde{P} \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} |T^n(\tilde{P})| \cdot |\mathcal{P}_n(\mathcal{X})|$$

$$\leq \max_{\tilde{P} \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} |T^n(\tilde{P})| \cdot (n + 1)^{|\mathcal{X}|}$$

$$\leq \max_{\tilde{P} \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} e^{nH(\tilde{P})} \cdot (n + 1)^{|\mathcal{X}|}$$

$$= \exp \left( n \max_{\tilde{P} \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} H(\tilde{P}) \right) \cdot (n + 1)^{|\mathcal{X}|}$$

$$= (n + 1)^{|\mathcal{X}|} e^{nH(P^*)},$$

where in (7) we have used TT1; and in (8) we used the upper bound of TT3. Since

$$H(P^*) = \max_{P \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} H(P) \leq \sup_{Q \in \mathcal{F}} H(Q),$$

the upper bound follows.

For the lower bound, suppose that we can find a sequence $\{P_n \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})\}$, such that

$$\lim_{n \to \infty} H(P_n) = \sup_{Q \in \mathcal{F}} H(Q).$$

(11)

We now proceed as follows:

$$|T^n(\mathcal{F})| = |T^n(\mathcal{F} \cap \mathcal{P}_n(\mathcal{X}))|$$

$$= \sum_{P \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} |T^n(P)|$$

$$\geq \sum_{P \in \mathcal{F} \cap \mathcal{P}_n(\mathcal{X})} \frac{1}{(n + 1)^{|\mathcal{X}|}} e^{nH(P)}$$

$$\geq \frac{1}{(n + 1)^{|\mathcal{X}|}} e^{nH(P_n)},$$

where (14) follows from the lower bound of TT3, and where in (15) we drop all elements of the sum apart from one. Actually, we pick one from the sequence of types defined in (11). Taking the logarithm and the limit then yields:

$$\lim_{n \to \infty} \frac{1}{n} \log |T^n(\mathcal{F})| \geq \lim_{n \to \infty} \left\{ \frac{|\mathcal{X}| \log(n + 1)}{n} + H(P_n) \right\}$$

(16)

$$= \lim_{n \to \infty} H(P_n)$$

(17)
Here, (16) follows from (15); and the last equality follows by assumption (11). This proves the second statement. The final statement now follows because from the first it follows that
\[
\lim_{n \to \infty} \frac{1}{n} \log \left| T^n(\mathcal{F}) \right| \leq \sup_{Q \in \mathcal{F}} H(Q).
\]

**Problem 3**

**Sanov’s Theorem for Bernoulli-Random Variables**

a) Using the fact that \( \bar{X}_n \) gives the ratio of 1’s in the sequence we get
\[
\Pr[\bar{X}_n \geq p] = \Pr(\text{all } n\text{-sequences with no less than } np \text{ 1’s})
\]
\[
= \Pr(\text{all } n\text{-sequences with no less than } \lceil np \rceil \text{ 1’s})
\]
\[
= \sum_{k=\lceil np \rceil}^{n} \Pr(\text{all } n\text{-sequences with } k \text{ 1’s})
\]
\[
= \sum_{k=\lceil np \rceil}^{n} \binom{n}{k} \Pr(\text{one } n\text{-sequence with } k \text{ 1’s})
\]
\[
= \sum_{k=\lceil np \rceil}^{n} \binom{n}{k} q^k (1-q)^{n-k}.
\]

b) Let’s compare two terms of the sum for \( k \) and \( k+1 \) and define the following ratio:
\[
f(k, n, q) \triangleq \frac{\binom{n}{k} q^k (1-q)^{n-k}}{\binom{n+1}{k+1} q^{k+1} (1-q)^{n-k-1}} = \frac{(1-q)(n-k-1)! (k+1)!}{q! (n-k)!} = \frac{(1-q)(k+1)}{q(n-k)}.
\]
We want to show that for all \( k \geq \lceil np \rceil \) this ratio is larger than 1, so that the term for \( k \) is larger than the term for \( k+1 \), i.e., the choice \( k = \lceil np \rceil \) leads to the largest term of the sum.
To that goal firstly compute the following derivative:
\[
\frac{\partial f(k, n, q)}{\partial k} = \frac{(1-q)q(n-k) + (1-q)(k+1)q}{q^2(n-k)^2} = \frac{(1-q)(n+1)}{q(n-k)^2} > 0.
\]
Hence, \( f \) is strictly increasing. So once \( f \) is larger than 1 it will always remain larger than 1.
It is simple to derive the (usually noninteger) value of \( k \) that achieves 1:
\[
f(k^*, n, q) = 1 \iff k^* = nq - 1 + q.
\]
Hence, it only remains to prove that \( k^* \leq \lceil np \rceil \). This follows because \( p > q \):
\[
k^* = nq - (1-q) \leq nq < np \leq \lceil np \rceil.
\]

c) We use the following bound from the script:
\[
\frac{1}{n+1} e^{nH_b(k/n)} \leq \binom{n}{k} \leq e^{nH_b(k/n)}.
\]
We will now show that the term for $k = \lceil np \rceil$ behaves approximately like $e^{-nD\ast}$ where

$$D\ast \triangleq D((p, 1-p) \parallel (q, 1-q)) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}.$$  

We will actually only derive an upper bound. A similar lower bound will be found in Part d).

$$\frac{1}{n} \log \left( \binom{n}{\lceil np \rceil} q^{\lceil np \rceil} (1-q)^{n-\lceil np \rceil} \right)$$

$$\leq \frac{1}{n} \log \left( e^{n \log p / n} \binom{n}{\lceil np \rceil} q^{\lceil np \rceil} (1-q)^{n-\lceil np \rceil} \right)$$

$$= H_b \left( \frac{\lceil np \rceil}{n} \right) + \frac{\lceil np \rceil}{n} \log q + \frac{n - \lceil np \rceil}{n} \log(1-q)$$

$$\leq H_b \left( \frac{np}{n} \right) + \left( \frac{np}{n} \right) \log q + \frac{n - (np + 1)}{n} \log(1-q)$$

$$= H_b(p) + p \log q + (1-p) \log(1-q) - \frac{1}{n} \log(1-q)$$

$$= -p \log p - (1-p) \log(1-p) + p \log q + (1-p) \log(1-q) - \frac{1}{n} \log(1-q)$$

$$= -p \log \frac{p}{q} - (1-p) \log \frac{1-p}{1-q} - \frac{1}{n} \log(1-q)$$

$$= -D\ast - \frac{1}{n} \log(1-q)$$

$$= -D\ast + \delta_1.$$  

Here, the two inequalities follow from (19) and from bounding $\lceil np \rceil$ by $np + 1$ or $np$, respectively. Note that $\log q$ and $\log(1-q)$ are negative and that because $p > q \geq \frac{1}{2}$ the binary entropy function $H_b(\cdot)$ is decreasing. The final equality should be understood as definition for $\delta_1$. Note that $\delta_1 \geq 0$ and tends to zero as $n \to \infty$.

**d)** Using all results above we get the following upper bound:

$$\frac{1}{n} \log \Pr[\bar{X}_n \geq p] = -\frac{1}{n} \log \left( \sum_{k=\lceil np \rceil}^{n} \binom{n}{k} q^k (1-q)^{n-k} \right)$$

$$\leq \frac{1}{n} \log \left( (n - \lceil np \rceil + 1) \binom{n}{\lceil np \rceil} q^{\lceil np \rceil} (1-q)^{n-\lceil np \rceil} \right)$$

$$= \frac{1}{n} \log (n - \lceil np \rceil + 1) + \frac{1}{n} \log \left( \binom{n}{\lceil np \rceil} q^{\lceil np \rceil} (1-q)^{n-\lceil np \rceil} \right)$$

$$\leq \frac{1}{n} \log (n - np + 1) - D\ast + \delta_1$$

$$= -D\ast + \delta_2$$

with

$$\delta_2 = \frac{1}{n} \log \frac{n(1-p) + 1}{1-q} \to 0 \quad \text{as } n \to \infty.$$  

Here (20) follows from a); (21) from b) by upper-bounding every term in the sum by the first (which is the largest); and (23) from c).

Similarly, we can derive a lower bound:

$$\Pr[\bar{X}_n \geq p] = \sum_{k=\lceil np \rceil}^{n} \binom{n}{k} q^k (1-q)^{n-k}.$$
Problem 4

Sanov’s Theorem and Conditional Limit
Theorem for Gaussian RVs

a) Note that the squared sum of two independent Gaussian random variables $X_1^2 + X_2^2$ is exponentially distributed with mean $\mathbb{E}[X_1^2 + X_2^2] = 2\sigma^2$. Hence, assume $n$ to be even and let $\ell = \frac{n}{2}$. Furthermore, let

\[
Y = \sum_{k=1}^\ell (X_{2k-1}^2 + X_{2k}^2).
\]

Hence, for $\ell = 1$, $Y \sim \mathcal{E}xp(2\sigma^2)$.

Next, we want to figure out the distribution of $Y$ for $\ell > 1$. For $\ell = 2$ we get the sum of two independent exponentially distributed RVs, which leads to the convolution of two exponential distributions:

\[
f_Y(y) = \int_0^y \frac{1}{2\sigma^2} e^{-\frac{y-t}{2\sigma^2}} \cdot \frac{1}{2\sigma^2} e^{-\frac{t}{2\sigma^2}} \, dt = \frac{t}{(2\sigma^2)^2} e^{-\frac{t}{2\sigma^2}} \quad \text{for } \ell = 2.
\]
Continuing in similar fashion we get for general $\ell$:

$$f_Y(y) = \frac{y^{\ell-1} e^{-\frac{y^2}{2\sigma^2}}}{(2\sigma^2)^\ell (\ell - 1)!}.$$  

Hence,

$$\Pr \left[ \frac{1}{n} \sum_{k=1}^{n} X_k^2 \geq \alpha^2 \right] = \Pr \left[ \sum_{j=1}^{\ell} \left( X_{2j-1}^2 + X_{2j}^2 \right) \geq n\alpha^2 \right]$$

$$= \Pr \left[ Y \geq 2\ell \alpha^2 \right]$$

$$= \int_{2\ell \alpha^2}^{\infty} \frac{y^{\ell-1} e^{-\frac{y^2}{2\sigma^2}}}{(2\sigma^2)^\ell (\ell - 1)!} \, dy$$

$$= \frac{1}{(2\sigma^2)^\ell (\ell - 1)!} \frac{(2\sigma^2)^\ell \Gamma \left( \ell, \frac{2\ell \alpha^2}{2\sigma^2} \right)}{\Gamma(\ell)}.$$  

To find the exponent we now need to derive the following limit:

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr \left[ \frac{1}{n} \sum_{k=1}^{n} X_k^2 \geq \alpha^2 \right] = \lim_{\ell \to \infty} \frac{1}{2\ell} \log \Pr \left[ Y \geq 2\ell \alpha^2 \right]$$

$$= \lim_{\ell \to \infty} \frac{1}{2\ell} \log \frac{\Gamma \left( \ell, \frac{\ell \alpha^2}{\sigma^2} \right)}{\Gamma(\ell)}$$

$$= \lim_{\ell \to \infty} \frac{1}{2\ell} \left( (\ell - 1) \log \left( \frac{\ell \alpha^2}{\sigma^2} \right) - \frac{\ell \alpha^2}{\sigma^2} \right)$$

$$- \ell \log \ell + \frac{1}{2} \log \ell - \frac{1}{2} \log 2\pi$$

$$= \frac{1}{2} \log \frac{\alpha^2}{\sigma^2} - \frac{\alpha^2}{2\sigma^2} + \frac{1}{2}. \quad (25)$$

Here we used the asymptotic expansion of the Gamma function and the incomplete Gamma function:

$$\log \Gamma(n) = n \log n - n - \frac{1}{2} \log 2\pi + o(n^{-1});$$

$$\Gamma(n, x) = x^{n-1} e^{-x} \left( 1 + O \left( \frac{1}{|x|} \right) \right).$$

b) In order to derive this result using Sanov’s theorem, we define

$$\mathcal{F} \triangleq \left\{ \tilde{Q} \in \mathcal{P}(\mathbb{R}) : \int_{-\infty}^{\infty} x^2 \, d\tilde{Q}(x) \geq \alpha^2 \right\}.$$  

Note that $\mathcal{F}$ is a “nice” set so that Sanov’s theorem says the following:

$$\lim_{n \to \infty} \frac{1}{n} \log \Pr \left[ \frac{1}{n} \sum_{k=1}^{n} X_k^2 \geq \alpha^2 \right] = - \inf_{\tilde{Q} \in \mathcal{F}} D(\tilde{Q} \| Q).$$
In order to compute this minimization, let \( \tilde{f}(x) \) denote the density corresponding to \( \tilde{Q} \). We use a Lagrange multiplier approach:

\[
L(\tilde{f}) = \int_{-\infty}^{\infty} \tilde{f}(x) \log \frac{\tilde{f}(x)}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \, dx + \lambda \left( \int_{-\infty}^{\infty} x^2 \tilde{f}(x) \, dx - \alpha^2 \right) + \mu \left( \int_{-\infty}^{\infty} \tilde{f}(x) \, dx - 1 \right)
\]

\[
= \int_{-\infty}^{\infty} \tilde{f}(x) \log \frac{\tilde{f}(x)}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \, dx + \frac{1}{2} \log 2\pi\sigma^2 + \left( \lambda + \frac{1}{2\sigma^2} \right) \int_{-\infty}^{\infty} x^2 \tilde{f}(x) \, dx - \lambda \alpha^2
\]

\[
+ \mu \left( \int_{-\infty}^{\infty} \tilde{f}(x) \, dx - 1 \right);
\]

\[
\frac{\partial L(\tilde{f})}{\partial \tilde{f}(x)} = \log \tilde{f}(x) + 1 + \left( \lambda + \frac{1}{2\sigma^2} \right) x^2 + \mu \quad \Rightarrow \quad \tilde{f}^*(x) = e^{\tilde{\mu} + \tilde{\lambda} x^2}.
\]

This is again a zero-mean Gaussian distribution! Since \( E_{\tilde{Q}^*}[X^2] \geq \alpha^2 \), it follows that \( \tilde{Q}^* = \mathcal{N}(0, \alpha^2) \).

Next we compute \( D(\tilde{Q}^* \parallel Q) \):

\[
- D(\tilde{Q}^* \parallel Q) = - \int_{-\infty}^{\infty} \tilde{f}^*(x) \log \frac{\tilde{f}^*(x)}{f(x)} \, dx
\]

\[
= \frac{1}{2} \log 2\pi\sigma^2 - \frac{1}{2} \log 2\pi\alpha^2 - \frac{1}{2\sigma^2} E_{\tilde{Q}^*}[X^2]
\]

\[
= \frac{1}{2} \log \frac{\alpha^2}{\sigma^2} + \frac{1}{2} - \frac{\alpha^2}{2\sigma^2},
\]

which is identical to (25).

c) We know from the conditional limit theorem that conditional on

\[
\frac{1}{n} \sum_{k=1}^{n} X_k^2 \geq \alpha^2
\]

the first couple of realizations of \( X_1, X_2, \ldots, X_m \) look like as if they were generated independently according to \( \tilde{Q}^* \) derived in b), i.e., they look like IID \( \sim \mathcal{N}(0, \alpha^2) \).

© Stefan M. Moser, 2015 8