Statistics G8243
Tuesday, March 31, 2009
Handout #13

Outline of solutions for HW Set # 3

1. —

2. *Strong converse in data compression.* This should be 80% identical to the proof of the converse part of the "error exponents" theorem from class.

3. *Counting.* We wish to count the number of sequences satisfying a certain property. Instead of directly counting the sequences, we will calculate the probability of the set under the uniform distribution. Since the uniform distribution puts a probability of $\frac{1}{m^n}$ on every sequence of length n, we can count the sequences by multiplying the probability of the set by m^n.

   The probability of the set can be calculated easily from Sanov's theorem. Let Q be the uniform distribution, and let E be the set of sequences of length n satisfying $\frac{1}{n}\sum g(x_i) \geq \alpha$. Then by Sanov's theorem, we have

   $$Q^n(E) \approx 2^{-nD(P^*\|Q)},$$

   where P^* is the distribution on A hat is closest to Q. Since Q is the uniform distribution, $D(P\|Q) = \log m - H(P)$, and therefore P^* is the distribution that has maximum entropy. Therefore, if we let,

   $$H^* = \max_{P:\sum_{i=1}^{m} P(i)g(i)\geq\alpha} H(P),$$

   we have

   $$Q^n(E) \approx 2^{-n(\log m - H^*)}.$$

   Multiplying this by m^n to find the number of sequences in this set, we obtain,

   $$|E| \approx 2^{-n\log m} 2^{nH^*} m^n = 2^{nH^*},$$

   where, as usual, the approximation is accurate to first order in the exponent.

4. *Large deviations.* Suppose $X_1, X_2, \ldots$ are IID random variables with distribution $Q(k) = p^{k-1}(1-p)$, for $k = 1, 2, \ldots$. Although the results we proved in class only apply to finite-valued random variables, we will ignore this issue here. Note, also, that the geometic distribution is defined in terms of the failure probability p here, so its in terms of $(1-p)$ in the more common notation. Also, we'll assume that α is larger than $1/(1-p)$, the mean of Q, so that the event we're looking at is indeed a large deviations event.

   (a) $\Pr\{\frac{1}{n}\sum_{i=1}^{n} X_i \geq \alpha\}$ can be evaluated, to first order in the exponent, by Sanov's theorem. If we let E be the set of distributions P on A such that $E_P(X) \geq \alpha$, then Sanov's theorem says that, to first order in the exponent,

   $$\Pr\left\{\frac{1}{n}\sum_{i=1}^{n} X_i \geq \alpha\right\} = Q^n\{\hat{P}_{X_1^n} \in E\} \approx 2^{-nD(P^*\|Q)},$$

where, $D(P^*\|Q) = \inf_{P\in E} D(P\|Q)$. From the "Gibbs' distributions" proposition we saw in class, it is easy to see that P^* will also be geometric here. Moreover, in order to satisfy the constraint $E_{P^*}(X) \geq \alpha$ with equality, we obtain that $P^*(k) = r^{k-1}(1-r)$, with $r = 1 - 1/\alpha$. Direct evaluation gives,

$$D(P^*\|Q) = -\log(\alpha(1-p)) + (\alpha - 1)\log\left(\frac{\alpha - 1}{\alpha p}\right).$$

(b) As in the previous part, using the conditional limit theorem, we get that the required conditional probability is $\approx P^*(k)$, with P^* as above.

(c) In this case, $r = 3/4$ and $D(P^*\|Q) = 3\log(3/2) - 1 \approx 0.7549$.

5. *The running problem.*
Let Q be the distribution of $X_i - Y_i$, so that $Q(z) = \sum_x Q_1(x)Q_2(x-z)$. Then just use Sanov's theorem as in the previous problem.

6. *Bent coins.* Let $\{X_i\}$ be IID $\sim Q$ where Q is the Binomial(m,q) distribution. This is very similar to problem 4. From the conditional limit theorem, as $n \to \infty$,

$$\Pr\left\{X_1 = k \;\middle|\; \frac{1}{n}\sum_{i=1}^n X_i \geq \alpha\right\} = Q^n(\hat{P}_{X_1^n} \in E) \to P^*(k),$$

where $E = \{P \in \mathcal{P} : E_P(X) \geq \alpha\}$, and P^* achives, $\min_{P\in E} D(P\|Q)$. From the "Gibbs' distributions" proposition we saw in class, it is easy to see that P^* will also be a binomial of the form Bin(m,λ). And in order to satisfy the constraint $E_{P^*}(X) = m\lambda \geq \alpha$ with equality, we obtain that P^* is Bin$(m, \alpha/m)$.

Note that, as in problem 4, this computation is only valid (and interesting) if α is strictly greater than the mean mq of Q.

7. *Large deviations below the mean.* Let $\{X_n\}$ be IID random variables with distribution Q on a finite alphabet A. For $\min\{A\} < c < E_Q[X]$, the upper bound of Cramèr's theorem here states that,

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^n X_i \leq c\right\} \leq \exp\{-n\Lambda^*(c)\},$$

where $\Lambda^*(c) = \sup_{\lambda\leq 0}[\lambda c - \Lambda(\lambda)]$, and the log-moment generating function $\Lambda(\lambda)$ is the same as before. The corresponding lower bound says that,

$$\liminf_{n\to\infty} \frac{1}{n}\ln\Pr\left\{\frac{1}{n}\sum_{i=1}^n X_i \leq c\right\} = -\Lambda^*(c).$$

The relevant properties of $\Lambda(\lambda)$ are that $\Lambda'(0) = E_Q(X)$, $\lim_{\lambda\to-\infty}\Lambda'(\lambda) = \min(A)$, $\Lambda''(\lambda) \geq 0$ for all $\lambda \leq 0$, and for each c in the given range, there exists a λ^* s.t. $\Lambda'(\lambda^*) = c$ and $\Lambda^*(c) = \lambda^* c - \Lambda(\lambda^*)$.

8. *Simulating rare events.* This is very much like problems 4 and 6; use the conditional limit theorem with Q being the uniform distribution on $\{1, 2, \ldots, 6\}$ and $n = 10,000$. We're interested in the event that $\frac{1}{n}\sum X_i \geq 1/4$, and, conditional on this event, by the conditional limit theorem the rolls will look like they were produced from P^*, where, $P^*(k)$ will be proportional to β^k for some $\beta > 0$, which is chosen so that $E_{P^*}(X) = 1/4$.

   (a) The analytical answer is the one given above.

   (b) If you want to simulate, simulate IID from P^*.

9. —