

ECE 520.447
Information Theory and Coding

Sanjeev Khudanpur

December 2-4, 2002

Sanov's Theorem for a Finite Alphabet¹

Using the notion of types, we can prove the following very useful result about the large deviations of the empirical mean of an i.i.d. process taking values in a discrete and finite set $\mathcal{X} = \{1, \dots, d\}$.

Sanov's Theorem: Let $\{X_t, \}_{t=1}^{\infty}$ be an i.i.d. process taking values in \mathcal{X} , with the common probability mass function Q . For any $G \subset \mathbb{P}^d$, the Q -probability of all sequences $x_1^n \in \mathcal{X}^n$ such that the corresponding type $P_{x_1^n}$ is in G , is

$$Q(P_{X_1^n} \in G) \leq (n+1)^d 2^{-nD(P^*||Q)}, \quad n = 1, 2, \dots,$$

where $D(P^*||Q) = \inf_{P \in G} D(P||Q)$. Further, if G is the closure of its interior, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log Q(P_{X_1^n} \in G) = -D(P^*||Q).$$

Proof: For a given n , observe that $P_{X_1^n}$ takes values only on \mathcal{T}^n , the set of all *types* with denominator n . We can therefore compute the probability of G by summing over all the types in G , and we already have an upper bound on the probability of each type class. Thus

$$\begin{aligned} Q(P_{X_1^n} \in G) &= Q(\mathcal{T}^n \cap G) = \sum_{P \in \mathcal{T}^n \cap G} Q(T_P^n) \\ &\leq \sum_{P \in \mathcal{T}^n \cap G} 2^{-nD(P||Q)} \\ &\leq \sum_{P \in \mathcal{T}^n \cap G} \sup_{P' \in \mathcal{T}^n \cap G} 2^{-nD(P'||Q)} \\ &\leq |\mathcal{T}^n \cap G| 2^{-n \inf_{P' \in G} D(P'||Q)}, \end{aligned}$$

¹This section and the next two sections on the Conditional Limit Theorem and Stein's Lemma have been adapted from T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

and the first assertion of the theorem follows from the cardinality bound for \mathcal{T}^n . Next, if G is the closure of its interior then it is closed and $D(P^*||Q)$ is attained by some $P^* \in G$ (why?). Further, any open (Euclidean) ball $B(P^*, \epsilon)$ of radius $\epsilon > 0$ containing P^* has a nonempty intersection with the interior of G . As n grows, \mathcal{T}^n becomes finer in its coverage of \mathbb{P}^d and thus we can find a type $P_n \in \mathcal{T}^n$ which is in the intersection of $B(P^*, \epsilon)$ and (the interior of) G . Therefore, *we can pick a sequence $P_n \in \mathcal{T}^n \cap G$ such that $D(P_n||Q) \rightarrow D(P^*||Q)$* . For this choice of P_n , we get

$$\begin{aligned} Q(P_{X_1^n} \in G) &= \sum_{P \in \mathcal{T}^n \cap G} Q(T_P^n) \\ &\geq Q(T_{P_n}^n) \\ &\geq (n+1)^{-d} 2^{-nD(P_n||Q)}, \end{aligned}$$

and therefore

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log Q(P_{X_1^n} \in G) &\geq \liminf_{n \rightarrow \infty} \left(-d \frac{\log(n+1)}{n} - D(P_n||Q) \right) \\ &= -D(P^*||Q). \end{aligned}$$

The second assertion follows by noting from the first assertion that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log Q(P_{X_1^n} \in G) \leq -D(P^*||Q).$$

Thus the limit exists and is the negative minimum divergence. ■

Homework: Note that the infimum in the definition of $D(P^*||Q)$ need not be attained by any P^* in G for the upper bound on $Q(G)$ to hold. To obtain a lower bound, we can bound the sum of $Q(T_P^n)$ below by the single largest term which, in turn, corresponds to the type with the smallest I-divergence from Q . However, to obtain the same exponent as in the upper bound, it becomes necessary to have that the minimum² I-divergence over all types in $\mathcal{T}^n \cap G$ goes to the infimum over probability measures in G .

1. Give an example of a set G for which this holds, but which does not meet the condition of the theorem.
2. Modify the arguments in the proof above to show that the theorem holds for this G .

Sanov's theorem holds in a more general setting than the one in which we we have proved it. In particular, the finiteness of the alphabet is not necessary. However, an intuitive understanding of "what goes on when a large deviation occurs" is best gleaned from this case.

²Why is the minimum guaranteed to exist?

A Conditional Limit Theorem

So far, we have been concerned with the probability of a large deviation of the empirical measure. The next result, which is frequently called a conditional limit theorem, answers the question, “Given that a large deviation of the empirical measure has occurred ($P_{X_1^n} \in G$), what can we say about $P_{X_1^n}$?” The qualitative answer is that if a large deviation has indeed occurred then, with high probability, $P_{X_1^n}$ is close to P^* , the measure in G which is closest to Q in I-divergence. Specifically, if we treat $P_{X_1^n}$ as a random vector taking values in \mathbb{P}^d (with the appropriate conditional measure), then $P_{X_1^n}$ converges to P^* in probability.

Conditional Limit Theorem: Let G be a closed, convex subset of \mathbb{P}^d with a nonempty interior and let $Q \notin G$ be the common probability measure for the i.i.d. sequence $\{X_t\}_{t=1}^\infty$. Let P^* be the unique minimizer of $D(P\|Q)$ over $P \in G$. Then, for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} Q \left(|P_{X_1^n}(i) - P^*(i)| > \epsilon \mid P_{X_1^n} \in G \right) = 0, \quad i = 1, \dots, d,$$

i.e. $P_{X_1^n}$ converges, conditionally, to P^* in probability.

Proof: Let $D^* = D(P^*\|Q)$, and for some $\delta > 0$ to be specified later, define

$$\begin{aligned} A &= \{P \in G : D(P\|Q) \leq D^* + \delta\} \\ B &= \{P \in G : D(P\|Q) > D^* + 2\delta\}. \end{aligned}$$

Now,

$$\begin{aligned} Q(P_{X_1^n} \in B) &= \sum_{P \in \mathcal{T}^n \cap B} Q(T_P^n) \\ &\leq \sum_{P \in \mathcal{T}^n \cap B} 2^{-nD(P\|Q)} \\ &\leq \sum_{P \in \mathcal{T}^n \cap B} 2^{-n(D^* + 2\delta)} \\ &\leq (n+1)^d 2^{-n(D^* + 2\delta)}. \end{aligned}$$

On the other hand

$$\begin{aligned} Q(P_{X_1^n} \in A) &= \sum_{P \in \mathcal{T}^n \cap A} Q(T_P^n) \\ &\geq \sum_{P \in \mathcal{T}^n \cap A} (n+1)^{-d} 2^{-nD(P\|Q)} \\ &\geq (n+1)^{-d} 2^{-n \min_{P \in \mathcal{T}^n \cap A} D(P\|Q)}. \end{aligned}$$

Since the interior of G is not empty and $D(P^*\|Q) = D^*$, there are *bona fide* types in A for all sufficiently large n . Therefore the minimum in the exponent above exists³ and is (obviously)

³Otherwise the minimum in an empty set is either undefined or (by convention) $+\infty$.

no greater than $D^* + \delta$. Therefore

$$Q(P_{X_1^n} \in A) \geq (n+1)^{-d} 2^{-n(D^*+\delta)}.$$

We can use these bounds on the probabilities of A and B together with Baye's rule to get

$$\begin{aligned} Q(P_{X_1^n} \in B \mid P_{X_1^n} \in G) &= \frac{Q(P_{X_1^n} \in B \cap G)}{Q(P_{X_1^n} \in G)} \\ &\leq \frac{Q(P_{X_1^n} \in B)}{Q(P_{X_1^n} \in A)} \\ &\leq \frac{(n+1)^d 2^{-n(D^*+2\delta)}}{(n+1)^{-d} 2^{-n(D^*+\delta)}} \\ &= (n+1)^{2d} 2^{-n\delta}. \end{aligned}$$

Therefore,

$$\lim_{n \rightarrow \infty} Q(P_{X_1^n} \in B \mid P_{X_1^n} \in G) = 0.$$

Now, note that

$$B' = G \setminus B = \{P \in G : D(P\|Q) \leq D^* + 2\delta\},$$

is also a closed, convex set. Since $P^* \in B'$, we can invoke the Pythagorean inequality for the I-projection and write

$$D(P\|P^*) + D(P^*\|Q) \leq D(P\|Q) \leq D^* + 2\delta, \quad P \in B',$$

and therefore $D(P\|P^*) \leq 2\delta$ for all $P \in B'$. To complete the proof of the theorem, observe that by Pinsker's inequality, given any $\epsilon > 0$ there is a $\delta = \delta(\epsilon, P^*)$ such that

$$D(P\|P^*) \leq 2\delta \implies \|P - P^*\|_1 \leq \epsilon \implies |P(i) - P^*(i)| \leq \epsilon, \quad i = 1, \dots, d.$$

For any given ϵ , if we choose our δ in this manner, it follows for every $i \in \mathcal{X}$ that

$$Q(|P_{X_1^n}(i) - P^*(i)| \leq \epsilon \mid P_{X_1^n} \in G) \geq Q(P_{X_1^n} \in B' \mid P_{X_1^n} \in G),$$

and we have shown that this probability goes to one. ■

Remark: We have proved a ‘‘marginal’’ version of the conditional limit theorem, which states that the empirical measure converges conditionally in probability to P^* . A stronger version holds, which states that the empirical measure of k -blocks, for ever fixed k , converges to the product measure $P^* \times \dots \times P^*$ on \mathcal{X}^k . Conditioned on a large deviation, therefore, the sequence X_1, \dots, X_n appears, with high likelihood, as if it were an i.i.d. sequence generated by the measure P^* . This forms the basis for techniques to quickly simulate rare events.

Hypothesis Testing: Stein's and Chernoff's Theorems⁴

Results with the flavor of a LDP are well known in the classical problem of binary hypothesis testing. Let the \mathcal{X} -valued observations X_1, \dots, X_n be i.i.d. with a common distribution which is one of two known alternatives P_0 and P_1 , and let $d : \mathcal{X}^n \rightarrow \{0, 1\}$ be a decision rule used to determine which of the two distributions generated the data. Define, as usual, the false alarm probability of the test and the probability of a miss as

$$\alpha_n^d = \alpha(A_n^d) = P_0(A_n^d) \quad \text{where} \quad A_n^d = \{x_1^n \in \mathcal{X}^n : d(x_1^n) = 1\},$$

and $\beta_n^d = \beta(N_n^d) = P_1(N_n^d) \quad \text{where} \quad N_n^d = \mathcal{X}^n \setminus A_n^d,$

respectively, where A designates the region where the alternate hypothesis P_1 is accepted, and N , the region where the null hypothesis P_0 is determined to hold. We have already seen that if A_n^d is designed so as to *exclude* all the types "close" to P_0 , then α_n^d goes to zero exponentially fast. On the other hand, if A_n^d *includes* all the types close to P_1 , then β_n^d goes to zero exponentially fast as well. We shall now state and prove two results which indicate the best exponents with which these probabilities can be made to go to zero.

Best Exponent for β_n when α_n is Fixed (Howsoever Small)

First consider a formulation similar to the standard Neyman-Pearson test in which we seek a test d^* which minimizes β_n^d among all tests for which $\alpha_n^d < \epsilon$ for some fixed $\epsilon > 0$. To minimize β_n , the optimal test must include in $A_n^{d^*}$ as many types close to P_1 as possible without including P_0 . Note that since the P_0 -probability of $N_n^{d^*}$ is not required to decay exponentially, it may be possible to let $A_n^{d^*}$ include types closer and closer to P_0 as n increases, without violating $\alpha_n < \epsilon$, as long as the null hypothesis P_0 is always excluded from $A_n^{d^*}$.

Stein's Lemma: Let $\{X_t\}_{t=1}^\infty$ be an i.i.d. sequence taking values in \mathcal{X} and having the common measure Q . Let P_0 and P_1 be probability mass functions on \mathcal{X} such that $D(P_0||P_1) < +\infty$. For the hypothesis test between the alternatives $Q = P_0$ and $Q = P_1$, let

$$\beta_n(\epsilon) = \min_{d: \alpha_n^d \leq \epsilon} \beta_n^d.$$

Then

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n(\epsilon) = -D(P_0||P_1).$$

⁴Adapted from T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

Proof: The proof of the lemma is in three steps. First we construct a sequence of tests for which $\alpha_n^d \rightarrow 0$. Therefore our tests are candidates in the minimization which defines $\beta_n(\epsilon)$ for all sufficiently large n for every $\epsilon > 0$. Then we show that for this sequence of tests, β_n^d decays with the exponent in the assertion of the lemma. Finally, we show that no other sequence of tests d' can have a better exponent of decay for $\beta_n^{d'}$.

For any $\delta > 0$, define a sequence of tests d using the decision regions

$$N_n^d = \left\{ x_1^n \in \mathcal{X}^n : 2^{n(D(P_0\|P_1)-\delta)} \leq \frac{P_0(x_1^n)}{P_1(x_1^n)} \leq 2^{n(D(P_0\|P_1)+\delta)} \right\}, \quad n = 1, 2, \dots,$$

and note that under the null hypothesis, since $E_{P_0} \left[\log \frac{P_0(X)}{P_1(X)} \right] = D(P_0\|P_1)$,

$$P_0(N_n^d) = P_0 \left(D(P_0\|P_1) - \delta \leq \frac{1}{n} \sum_{t=1}^n \log \frac{P_0(X_t)}{P_1(X_t)} \leq D(P_0\|P_1) + \delta \right) \rightarrow 1,$$

for every $\delta > 0$, by the law of large numbers. Therefore $P_0(A_n^d) \rightarrow 0$.

Next, note that by the construction of N_n^δ

$$\begin{aligned} \beta_n^d = P_1(N_n^\delta) &= \sum_{x_1^n \in N_n^\delta} P_1(x_1^n) \\ &\leq \sum_{x_1^n \in N_n^\delta} 2^{-n(D(P_0\|P_1)-\delta)} P_0(x_1^n) \\ &= 2^{-n(D(P_0\|P_1)-\delta)} P_0(N_n^\delta), \end{aligned}$$

giving us

$$\frac{1}{n} \log \beta_n^d \leq -D(P_0\|P_1) + \delta + \frac{\log(1 - \alpha_n^d)}{n}.$$

Similarly

$$\begin{aligned} \beta_n^d = P_1(N_n^\delta) &= \sum_{x_1^n \in N_n^\delta} P_1(x_1^n) \\ &\geq \sum_{x_1^n \in N_n^\delta} 2^{-n(D(P_0\|P_1)+\delta)} P_0(x_1^n) \\ &= 2^{-n(D(P_0\|P_1)+\delta)} P_0(N_n^\delta), \end{aligned}$$

which yields

$$\frac{1}{n} \log \beta_n^d \geq -D(P_0\|P_1) - \delta + \frac{\log(1 - \alpha_n^d)}{n}.$$

But $\alpha_n^d \rightarrow 0$ and the choice of $\delta > 0$ is arbitrary. It holds therefore that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^d = -D(P_0\|P_1).$$

Since the limit does not depend on ϵ , it holds trivially that $\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^d = -D(P_0 \| P_1)$.

Finally, let d' be any other sequence of tests such that $\alpha_n^{d'} \rightarrow 0$. Then

$$\begin{aligned}
\beta_n^{d'} = P_1(N_n^{d'}) &\geq P_1(N_n^{d'} \cap N_n^d) \\
&= \sum_{x_1^n \in N_n^{d'} \cap N_n^d} P_1(x_1^n) \\
&\geq \sum_{x_1^n \in N_n^{d'} \cap N_n^d} 2^{-n(D(P_0 \| P_1) + \delta)} P_0(x_1^n) \\
&= 2^{-n(D(P_0 \| P_1) + \delta)} P_0(N_n^{d'} \cap N_n^d) \\
&\geq 2^{-n(D(P_0 \| P_1) + \delta)} (1 - P_0(A_n^{d'}) - P_0(A_n^d)) \\
&= (1 - \alpha_n^{d'} - \alpha_n^d) 2^{-n(D(P_0 \| P_1) + \delta)}.
\end{aligned}$$

Since $\alpha_n^d, \alpha_n^{d'} \rightarrow 0$ and the choice of $\delta > 0$ is arbitrary, it holds that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \beta_n^{d'} \geq -D(P_0 \| P_1),$$

and the assertion of the lemma follows by noting that this holds for every $\epsilon > 0$. ■

Homework: Why do we need consider as possible competitors only those tests for which $\alpha_n^{d'} \rightarrow 0$? What if $\lim_n \alpha_n^{d'} > 0$ or $\liminf_n \alpha_n^{d'} = 0$, but $\alpha_n^{d'}$ does not converge? (Hint: Observe that the lemma asserts the asymptotics only for $\epsilon \rightarrow 0$.)

A Bayesian Formulation and Chernoff Information

The Neyman-Pearson treatment of the hypothesis testing problem, while appropriate for several applications, lacks two components: the treatment of the probabilities of the two kinds of errors is asymmetric, and prior knowledge about the likelihoods of the two hypotheses cannot be used in the formulation. An alternative approach is to minimize the probability of error given by

$$P_e^d = \pi_0 \alpha_n^d + \pi_1 \beta_n^d,$$

where $\pi_0, \pi_1 \geq 0$ are the *a priori* probabilities of the two hypotheses, $\pi_0 + \pi_1 = 1$, and the probabilities of error α_n^d and β_n^d for a decision rule d are as defined earlier.

Chernoff's Theorem: Let $\{X_t\}_{t=1}^\infty$ be an i.i.d. sequence taking values in \mathcal{X} and having the common measure Q . Let P_0 and P_1 be probability mass functions on \mathcal{X} with prior probabilities π_0 and π_1 respectively. For the hypothesis test between the alternatives $Q = P_0$ and $Q = P_1$, let

$$D^* = \lim_{n \rightarrow \infty} \min_{A_n^d \subseteq \mathcal{X}^n} -\frac{1}{n} \log P_e^d,$$

denote the best achievable error exponent. Then

$$D^* = D(P_{\lambda^*} \| P_0) = D(P_{\lambda^*} \| P_1),$$

where λ^* is chosen such that

$$P_{\lambda}(x) = \frac{P_0^{\lambda}(x)P_1^{1-\lambda}(x)}{\sum_{x' \in \mathcal{X}} P_0^{\lambda}(x')P_1^{1-\lambda}(x')}, \quad x \in \mathcal{X}, \quad 0 \leq \lambda \leq 1$$

achieves the equality $D(P_{\lambda} \| P_0) = D(P_{\lambda} \| P_1)$.

Homework: Verify that D^* as defined above is equal to the usual definition of the *Chernoff Information* between P_0 and P_1 , given by

$$C(P_0, P_1) = \min_{0 \leq \lambda \leq 1} \log \sum_{x \in \mathcal{X}} P_0^{\lambda}(x)P_1^{1-\lambda}(x).$$

Proof: We know that the optimal test is a likelihood ratio test of the form

$$d(x_1^n) = \begin{cases} 1 & \text{if } \frac{P_1(x_1^n)}{P_0(x_1^n)} > \eta, \\ 0 & \text{otherwise.} \end{cases}$$

Note that for i.i.d. sequences, the following are equivalent

- $\frac{P_1(x_1^n)}{P_0(x_1^n)} > \eta_n.$
- $\prod_{t=1}^n \frac{P_1(x_t)}{P_0(x_t)} > \eta_n.$
- $\sum_{t=1}^n \log \frac{P_1(x_t)}{P_0(x_t)} > \log \eta_n.$
- $\sum_{x \in \mathcal{X}} P_{x_1^n}(x) \log \frac{P_1(x)}{P_0(x)} > \frac{1}{n} \log \eta_n.$
- $D(P_{x_1^n} \| P_0) - D(P_{x_1^n} \| P_1) > \zeta_n.$

Therefore, the optimal test is given by the partition of \mathcal{T}^n as

$$A_n^{d^*} = \left\{ x_1^n \in \mathcal{X}^n : D(P_{x_1^n} \| P_0) > D(P_{x_1^n} \| P_1) + \zeta_n \right\},$$

for an appropriate choice of ζ_n . It is also clear from Sanov's theorem that up to the first order in the exponent

$$\alpha_n^{d^*} \approx 2^{-n \min_{P \in A_n^{d^*}} D(P \| P_0)}, \quad \text{and} \quad \beta_n^{d^*} \approx 2^{-n \min_{P \in N_n^{d^*}} D(P \| P_1)}.$$

Precisely, the average probability of error is bounded as

$$P_e^{d^*} = \pi_0 \alpha_n^{d^*} + \pi_1 \beta_n^{d^*}$$

$$\begin{aligned}
&\geq \max \left\{ \pi_0 \alpha_n^{d^*}, \pi_1 \beta_n^{d^*} \right\} \\
&\geq \min \{ \pi_0, \pi_1 \} (n+1)^{-d} 2^{-n \min \left\{ \min_{P \in A_n^{d^*}} D(P \| P_0), \min_{P \in N_n^{d^*}} D(P \| P_1) \right\}}, \quad \text{and} \\
P_e^{d^*} &\leq 2 \max \left\{ \pi_0 \alpha_n^{d^*}, \pi_1 \beta_n^{d^*} \right\} \\
&\leq 2^{1-n \min \left\{ \min_{P \in A_n^{d^*}} D(P \| P_0), \min_{P \in N_n^{d^*}} D(P \| P_1) \right\}}.
\end{aligned}$$

Since we are interested in the best error exponent, we must select ζ_n so as to maximize the exponent above. We shall return to this optimal test shortly.

To minimize $D(P \| P_1)$ subject to $D(P \| P_0) - D(P \| P_1) \leq \zeta$ and $\sum_{x \in \mathcal{X}} P(x) = 1$, we use the method of Lagrange multipliers. We minimize w.r.t. P the function

$$\begin{aligned}
L(P, \lambda, \nu) &= D(P \| P_1) + \lambda (D(P \| P_0) - D(P \| P_1)) + \nu \sum_{x \in \mathcal{X}} P(x) \\
&= \lambda D(P \| P_0) + (1 - \lambda) D(P \| P_1) + \nu \sum_{x \in \mathcal{X}} P(x).
\end{aligned}$$

Differentiating with respect to $P(x)$ for $x \in \mathcal{X}$, and setting the partial derivative to zero, we get that

$$\lambda \left(1 + \log \frac{P(x)}{P_0(x)} \right) + (1 - \lambda) \left(1 + \log \frac{P(x)}{P_1(x)} \right) + \nu = 0.$$

We then solve for ν using the constraint $\sum_{x \in \mathcal{X}} P(x) = 1$, to get

$$P_\lambda(x) = \frac{P_0^\lambda(x) P_1^{(1-\lambda)}(x)}{\sum_{x' \in \mathcal{X}} P_0^\lambda(x') P_1^{(1-\lambda)}(x')},$$

and $\lambda = \lambda(\zeta)$ is obtained by solving $D(P_\lambda \| P_0) - D(P_\lambda \| P_1) = \zeta$. It is easy to see that the distribution P_λ also minimizes $D(P \| P_0)$ among all P for which $D(P \| P_0) - D(P \| P_1) \geq \zeta$. Finally, note that $D(P_\lambda \| P_0)$ is monotonically decreasing, while $D(P_\lambda \| P_1)$ is monotonically increasing as λ goes from 0 to 1. Therefore

$$\min \left\{ \min_{P: D(P \| P_0) - D(P \| P_1) \geq \zeta} D(P \| P_0), \min_{P: D(P \| P_0) - D(P \| P_1) \leq \zeta} D(P \| P_1) \right\}$$

has a maximum value of $D^* = D(P_{\lambda^*} \| P_0) = D(P_{\lambda^*} \| P_1)$, attained by $\zeta^* = 0$. Returning to the upper and lower bounds for $P_e^{d^*}$ note that

$$\begin{aligned}
\liminf_{n \rightarrow \infty} \frac{1}{n} \log P_e^{d^*} &\geq - \lim_{n \rightarrow \infty} \min \left\{ \min_{P \in A_n^{d^*}} D(P \| P_0), \min_{P \in N_n^{d^*}} D(P \| P_1) \right\} \\
\limsup_{n \rightarrow \infty} \frac{1}{n} \log P_e^{d^*} &\leq - \lim_{n \rightarrow \infty} \min \left\{ \min_{P \in A_n^{d^*}} D(P \| P_0), \min_{P \in N_n^{d^*}} D(P \| P_1) \right\}
\end{aligned}$$

and as shown above, the exponent attains the maximum value of D^* for some sequence $\zeta_n \rightarrow 0$. ■