Cramer's Theorem and Beyond

Let $X_i, 1 \leq i \leq n$ be iid random variables with common pdf p_X. For any $a > \mathbb{E}[X]$, we have the large-deviations bound

$$P\left[\sum_{i=1}^{n} X_i \geq na\right] \leq e^{-n[sa-\mu(s)]}, \quad s > 0, \tag{1}$$

where $\mu(s) = \ln \mathbb{E}[e^{sX}]$ is the c.g.f. of X. The best bound is obtained by choosing $s = s^*$ that maximizes $sa - \mu(s)$.

Cramer's theorem states that the upper bound of (1) is tight in the exponent:

$$\lim_{n\to\infty}\left(-\frac{1}{n}\ln P\left[\sum_{i=1}^{n} X_i \geq na\right]\right) = s^*a - \mu(s^*). \tag{2}$$

This can be proven by application of Sanov's theorem from Information Theory [1].

Can we now go beyond Cramer's theorem and derive a *precise asymptotic expansion* for $P[\sum_{i=1}^n X_i \geq na]$ as $n \to \infty$? The answer is yes. The result is the Bahadur-Rao theorem [2, 3] [4, Ch. 3.7]

$$P\left[\sum_{i=1}^{n} X_i \geq na\right] \sim \frac{e^{-n[s^*a-\mu(s^*)]}}{\sqrt{2\pi(s^*)^2\mu''(s^*)\,n}}, \tag{3}$$

where $\mu''(s)$ denotes the second derivative of the function $\mu(s)$. We have used the standard notation $f(n) \sim g(n)$ for asymptotic equality of two functions f and g as $n \to \infty$, i.e., $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$. Due to the term $n^{1/2}$ in the denominator of (3), the upper bound of (1) is increasingly **loose** as $n \to \infty$. (However, as discussed in class, this is not necessarily a concern, because an approximation that is tight in the exponent is usually satisfactory.)

Proof of (3). Let $n = 1$ and define the **tilted distribution**

$$p_{X_s}(x) = \frac{e^{sx}p(x)}{\int e^{sx}p(x)\,dx} = e^{sx-\mu(s)}p(x), \quad s > 0. \tag{4}$$

Observe that $p_{X_s}(x) = p(x)$ in the case $s = 0$. The first two cumulants of p_{X_s} are given by

$$\begin{aligned}
\mathbb{E}_{p_{X_s}}(X) &= \int x p_{X_s}(x)\,dx = \mu'(s), \\
\mathrm{Var}_{p_{X_s}}(X) &= \mu''(s).
\end{aligned}$$

(This proves that $\mu''(s) \geq 0$, hence that the c.g.f. is convex.)

Using the definition (4), we may write

$$\begin{aligned} P[X \geq a] &= \int_a^\infty p(x)\,dx \\ &= \int_a^\infty e^{-[sx-\mu(s)]}\,p_{X_s}(x)\,dx \\ &= e^{-[sa-\mu(s)]}\int_a^\infty e^{s(a-x)}\,p_{X_s}(x)\,dx. \end{aligned} \tag{5}$$

Observe that the value of the above integral is at most 1, hence the large-deviations upper bound $P[X \geq a] \leq e^{-[sa-\mu(s)]}$ follows directly and without explicit application of Markov's inequality.

Also observe that the integral in (5) represents the expectation of the function $e^{s(a-x)}\,1_{\{x\geq a\}}$ when the random variable X is distributed as p_{X_s}. Now we choose $s = s^*$ satisfying $\mu'(s^*) = a$. Define the normalized random variable

$$V = \frac{X - \mathbb{E}_{p_{X_s}}(X)}{\sqrt{\mathrm{Var}_{p_{X_s}}(X)}} = \frac{X-\mu'(s)}{\sqrt{\mu''(s)}} = \frac{X-a}{\sqrt{\mu''(s^*)}} \tag{6}$$

which has zero mean, unit variance, and pdf p_V. Using this notation and changing variables, we write (5) as

$$P[X \geq a] = e^{-[sa-\mu(s^*)]}\int_0^\infty e^{-s^*\sqrt{\mu''(s^*)}\,v}p_V(v)\,dv. \tag{7}$$

Now consider $n \geq 1$ and apply the above formulas with $\sum_{i=1}^n X_i$ in place of X. The c.g.f. for $\sum_{i=1}^n X_i$ is n times for c.g.f. for X. We obtain

$$P\left[\sum_{i=1}^n X_i \geq na\right] = e^{-n[sa-\mu(s^*)]}\int_0^\infty e^{-s^*\sqrt{\mu''(s^*)n}\,v}p_V(v)\,dv. \tag{8}$$

Moreover, as $n \to \infty$, it follows from the CLT that V converges in distribution to a normal distribution. While that convergence is slow in the tails of p_V, we now analyze (8) and show that only the value of p_V at $v = 0$ matters. Indeed let

$$\epsilon = \frac{1}{s^*\sqrt{\mu''(s^*)n}} \tag{9}$$

which tends to zero as $n^{-1/2}$. The integral in (7) is of the form

$$\int_0^\infty e^{-\frac{v}{\epsilon}}\,p_V(v)\,dv = \epsilon\int_0^\infty \underbrace{\left(\frac{1}{\epsilon}\,e^{-\frac{v}{\epsilon}}\right)}_{\text{``Dirac−like''}}\,p_V(v)\,dv \;\sim \epsilon\,p_V(0) \quad \text{as } \epsilon \to 0. \tag{10}$$

Combining (8), (9), (10), and the fact that $p_V(0) \to \frac{1}{\sqrt{2\pi}}$ as $n \to \infty$, we obtain the desired asymptotic equality

$$P\left[\sum_{i=1}^n X_i \geq na\right] \sim \frac{e^{-n[sa-\mu(s^*)]}}{\sqrt{2\pi(s^*)^2\mu''(s^*)n}} \quad \text{as } n \to \infty. \tag{11}$$

Chernoff Bounds for Detection of iid Gaussian Signals in iid Gaussian Noise

Problem. Consider the hypothesis test

$$\begin{cases} H_0: & Y_k = N_k \quad ,1 \le k \le n \\ H_1: & Y_k = S_k + N_k \quad ,1 \le k \le n \end{cases} \tag{12}$$

where $\{S_k\}$ are iid $\mathcal{N}(0, \sigma_s^2)$ and $\{N_k\}$ are iid $\mathcal{N}(0, \sigma^2)$; moreover $\{S_k\}$ and $\{N_k\}$ are independent.

Consider a loglikelihood ratio test with threshold τ of the form $n\overline{\tau}$, i.e., the threshold increases linearly with the number of data. For this test, the error exponents for P_F and P_M are given by

$$e_F = \lim_{n\to\infty} \left[-\frac{1}{n} \ln P_F\right] \tag{13}$$

$$e_M = \lim_{n\to\infty} \left[-\frac{1}{n} \ln P_M\right], \tag{14}$$

respectively. The problem is to derive expressions for both error exponents in terms of the normalized threshold $\overline{\tau}$. Of particular interest is the case $\overline{\tau} = 0$, which, as discussed in class, corresponds to the Bayesian case (the prior has no influence as $n \to \infty$.)

We will also specialize our results to the case of a low signal-to-noise ratio (per sample), to the case of a high SNR, and we will compare these results with those in the coherent-detection case.

Solution. To simplify the notation, define the signal-to-noise ratio per sample

$$\xi = \frac{\sigma_s^2}{\sigma^2}.$$

The hypothesis test is equivalent to testing

$$H_0 : Y_k \overset{iid}{\sim} \mathcal{N}(0,1) \quad \text{versus} \quad H_1 : Y_k \overset{iid}{\sim} \mathcal{N}(0, 1+\xi), \quad 1 \le k \le n.$$

To make the dependency of the error exponents on $\overline{\tau}$ and ξ clear, we write them as $e_F(\overline{\tau}, \xi)$ and $e_M(\overline{\tau}, \xi)$, respectively. The (per sample) cumulant-generating function

$$\mu(s) \triangleq \mathbb{E}_0[\ln L(Y)] = \frac{1}{2} \ln \frac{(\sigma^2)^s(\sigma^2 + \sigma_s^2)^{1-s}}{s\sigma^2 + (1-s)(\sigma^2 + \sigma_s^2)}, \quad 0 < s < 1,$$

takes the form

$$\begin{aligned} \mu(s, \xi) &= \frac{1}{2} \ln \frac{(1+\xi)^{1-s}}{s + (1-s)(1+\xi)} \\ &= \frac{1}{2} \ln \frac{(1+\xi)^{1-s}}{1 + (1-s)\xi} \\ &= \frac{1}{2}[(1-s)\ln(1+\xi) - \ln(1 + (1-s)\xi)]. \end{aligned} \tag{15}$$

From the theory covered in class we obtain

$$\begin{aligned} e_F(\overline{\tau},\xi) &= s^*\overline{\tau} - \mu(s^*,\xi) && (16) \\ e_M(\overline{\tau},\xi) &= -\overline{\tau} + e_F(\overline{\tau},\xi) && (17) \end{aligned}$$

where the optimal Chernoff exponent s^* satisfies the equation

$$\left.\frac{d}{ds}\mu(s,\xi)\right|_{s=s^*} = \overline{\tau}. \tag{18}$$

From (15) we obtain

$$\frac{d}{ds}\mu(s,\xi) = \frac{1}{2}\left[-\ln(1+\xi) + \frac{\xi}{1+(1-s)\xi}\right].$$

Substituting into (18), we obtain

$$1 + (1-s^*)\xi = \frac{\xi}{2\overline{\tau} + \ln(1+\xi)}$$

and therefore

$$s^* = -\frac{1}{2\overline{\tau} + \ln(1+\xi)} + \frac{1}{\xi} + 1.$$

Replacing in (16) and (17) and combining terms, we obtain the desired error exponents $e_F(\overline{\tau},\xi)$ and $e_M(\overline{\tau},\xi)$.

Range of $\overline{\tau}$. The equations above are valid when the normalized threshold satisfies

$$-D(p_0||p_1) < \overline{\tau} < D(p_1||p_0). \tag{19}$$

The two Kullback-Leibler divergences above are given by

$$\begin{aligned} -D(p_0||p_1) &= \left.\frac{d}{ds}\mu(s,\xi)\right|_{s=0} = \frac{1}{2}\left[-\ln(1+\xi) + \frac{\xi}{1+\xi}\right] = -\psi\left(\frac{1}{1+\xi}\right), && (20) \\ D(p_1||p_0) &= \left.\frac{d}{ds}\mu(s,\xi)\right|_{s=1} = \frac{1}{2}[-\ln(1+\xi) + \xi] = \psi(1+\xi). && (21) \end{aligned}$$

In (20), (21), we have used the function $\psi(x) = \frac{1}{2}(-1+x-\ln x)$, which is positive, unimodal, and achieves its minimum at $x = 1$. The value of the minimum is $\psi(1) = 0$.

Recall that for any $\overline{\tau}$ satisfying (19), we have $e_F(\overline{\tau},\xi) \le D(p_0||p_1)$ and $e_M(\overline{\tau},\xi) \le D(p_1||p_0)$.

Bayesian case. In the special case $\overline{\tau} = 0$, the two error exponents are equal:

$$e_F(0,\xi) = e_M(0,\xi) = \psi\left(\frac{\ln(1+\xi)}{\xi}\right).$$

Low SNR. Another special case of interest is the low signal-to-noise-ratio case, $\xi \ll 1$. Then the equations above simplify to

$$\mu(s,\xi) \sim -\frac{s(1-s)}{4}\xi \quad \text{as } \xi \to 0$$

and the optimal Chernoff exponent s^* tends to the Bhattacharyya exponent, $\frac{1}{2}$ (because the function $s(1-s)$ is maximized for $s = \frac{1}{2}$). Hence

$$e_F(0,\xi) = e_M(0,\xi) \sim -\mu\left(\frac{1}{2},\xi\right) \sim \frac{\xi^2}{16} \quad \text{as } \xi \to 0.$$

High SNR. We have $1 - s^* \sim \frac{1}{\ln \xi}$ and

$$e_F(0,\xi) = e_M(0,\xi) = \psi\left(\frac{\ln \xi}{\xi}\right) \sim \frac{1}{2}\ln \xi \quad \text{as } \xi \to \infty,$$

where in the last step we have used the asymptotic property $\psi(x) \sim \frac{1}{2}\ln\frac{1}{x}$ as $x \to 0$.

Coherent Case. It is interesting to compare that error exponent with the one for coherent detection of signals in iid Gaussian noise. To make a fair comparison with the stochastic-signal case, we assume the total energy of the known signal is $n\sigma_s^2$. For Bayesian detection with equal priors, we have

$$P_F^{\text{coh}}(\xi) = P_M^{\text{coh}}(\xi) = Q\left(\frac{\sqrt{n\xi}}{2}\right)$$

Since $\ln Q(x) \sim -\frac{x^2}{2}$ as $x \to \infty$, the error exponents for the coherent case are given by

$$e_F^{\text{coh}}(\xi) = e_M^{\text{coh}}(\xi) = \frac{\xi}{8}, \quad \forall \xi,$$

i.e., they are much better those in the stochastic case when $\xi \ll 1$ or $\xi \gg 1$!

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Ed., Wiley, 2006.

[2] D. Blackwell and J. L. Hodges, "The Probability in the Extreme Tail of a Convolution," *Ann. Math. Stat.*, Vol. 30, pp. 1113—1120, 1959.

[3] R. Bahadur and R. Ranga Rao, "On Deviations of the Sample Mean," *Ann. Math. Stat.*, Vol. 31, pp. 1015—1027, 1960.

[4] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, New York, 1998.