ASYMPTOTIC BEHAVIOUR OF POWERS
OF DICHOTOMIES

by

Erik N. Torgersen
Corrections to: ASYMPTOTIC BEHAVIOUR OF POWERS OF DICHOTOMIES.

Statistical research report No. 6, 1974 by Erik N. Togersen.

<table>
<thead>
<tr>
<th>page</th>
<th>line</th>
<th>old version</th>
<th>new version</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>7</td>
<td>Zolotarev's sharpening [ ]</td>
<td>Zolotarev's sharpening [12]</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
<td>Chernoff [ ]</td>
<td>Chernoff [3]</td>
</tr>
<tr>
<td>2.7</td>
<td>6</td>
<td>Put $A_w = \sum_{j=1}^{\infty} [</td>
<td>X_j</td>
</tr>
<tr>
<td>2.7</td>
<td>7</td>
<td>$\mathbb{P}(\bar{X}_n \geq a</td>
<td>A_w)$</td>
</tr>
<tr>
<td>2.7</td>
<td>9</td>
<td>$\mathbb{E}[e^{t(X_j-a)}</td>
<td>A_w]$</td>
</tr>
<tr>
<td>3.10</td>
<td>3</td>
<td>Let $\xi = (\lambda, \mu, \sigma, \rho)$</td>
<td>Let $\xi = (\lambda, \mu, \sigma, \rho)$</td>
</tr>
<tr>
<td>3.10</td>
<td>7</td>
<td>$\int \psi d \mathcal{K}$</td>
<td>$\int \psi d \mathcal{K}$</td>
</tr>
<tr>
<td>3.12</td>
<td>11</td>
<td>$\delta(\xi, \xi') = 0$</td>
<td>$\delta(\xi, \xi') = 0$</td>
</tr>
<tr>
<td>4.5</td>
<td>5</td>
<td>It follows that the convergence is uniform on $[0,1]$.</td>
<td>It follows that $[B_\lambda(\xi^n \sqrt{n})/L(t_o)^n$ converges, as $n \to \infty$, uniformly in $\lambda \in [0,1]$.</td>
</tr>
<tr>
<td>4.5</td>
<td>10</td>
<td>Efron and Truax [ ]</td>
<td>Efron and Truax [2] *</td>
</tr>
<tr>
<td>4.7</td>
<td>4</td>
<td>Chernoff [ ]</td>
<td>Chernoff [1] *</td>
</tr>
<tr>
<td>5.3</td>
<td>3</td>
<td>Joshi [ ]</td>
<td>Joshi [2] *</td>
</tr>
<tr>
<td>5.4</td>
<td>4</td>
<td>$\tau = \text{Var}_P \xi$</td>
<td>$\tau^2 = \text{Var}_P \xi$</td>
</tr>
<tr>
<td>5.7</td>
<td>1</td>
<td>Efron [ ]</td>
<td>Efron [1] *</td>
</tr>
<tr>
<td>5.7</td>
<td>5</td>
<td>proof in [ ]</td>
<td>proof in [1] *</td>
</tr>
<tr>
<td>6.1</td>
<td>4</td>
<td>in [ ]</td>
<td>in [2] *</td>
</tr>
<tr>
<td>6.1</td>
<td>5</td>
<td>Le Cam's paper [ ]</td>
<td>Le Cam's paper [2] *</td>
</tr>
<tr>
<td>6.3</td>
<td>7</td>
<td>Torgersen [ ]</td>
<td>Torgersen [3] *</td>
</tr>
</tbody>
</table>

*) The references are numbered according to the list of references appearing at the end of the section.
Content

PD 0 Summary

PD 1 Introduction

PD 2 Some facts on very large deviations

PD 3 Comparison of pseudo dichotomies
   in terms of functionals on the
   set of convex functions on [0,1]

PD 4 Asymptotic Baye's risk for testing
   problems

PD 5 Asymptotic behaviour of probabilities
   of errors of the first and the second
   kind

PD 6 Asymptotic behaviour of deficiencies

PD 7 References
Summary.

Consider random variables \(X, Y, \ldots\) whose distributions are known except for an unknown parameter \(\theta\) belonging to a known two-point set. Let \(X_1, X_2, \ldots\) and \(Y_1, Y_2, \ldots\) be independent observations of, respectively, \(X\) and \(Y\). How does the information yielded by \((X_1, X_2, \ldots, X_n)\) compare with the information yielded by \((Y_1, Y_2, \ldots, Y_n)\) when \(n\) is large?

Let \(\mathcal{M}_a\) and \(\mathcal{M}_i\) denote, respectively, a totally informative and a totally uninformative experiment. Furthermore denote by \(\Delta\) the distance between experiments introduced by LeCam 1964. Then, for any variable \(X\):

\[
1 - \Delta(X, \mathcal{M}_i) \leq \Delta(X, \mathcal{M}_a) < 2 \frac{1 - \Delta(X, \mathcal{M}_i) + \Delta(X, \mathcal{M}_a)}{\Delta(X, \mathcal{M}_a)}
\]

Combining this inequality with Chernoff's result on the exponential rate of asymptotic Baye's risk we find that

\[
\sqrt[n]{\Delta((X_1, X_2, \ldots, X_n), (Y_1, Y_2, \ldots, Y_n))} \leq \text{max}(c_X, c_Y)
\]

provided the experiments defined by \(X\) and \(Y\) are not equivalent. Here \(c_X(c_Y)\) denote the greatest lower bound of the Hellinger transform of \(X(Y)\).

In order to obtain inequalities for concave approximations to the kernel of the Hellinger transform, we generalized the sub linear function criterion as follows. Let \((x, \mathcal{X})\) and \((y, \mathcal{Y})\) be measurable spaces with, respectively, probability
measures $P_1, P_2$ and $Q_1, Q_2$. Suppose the dichotomy
\( \mathcal{G} = ((x, \mathcal{A}), (P_1, P_2)) \) is \((\epsilon_1, \epsilon_2)\) deficient w.r.t.
\( \mathcal{F} = ((y, \mathcal{B}), (Q_1, Q_2)) \). Then, for any convex function \( \varphi \) on
\([0,1]\),

\[
\epsilon_1 [\varphi'(1) - (\varphi(1) - \varphi(0))] + \epsilon_2 [(\varphi(1) - \varphi(0)) - \varphi'(0)] \geq 4 \int \varphi d(T - S)
\]

where \( S \) is the distribution of \( dP_2/d(P_1 + P_2) \) w.r.t \((P_1 + P_2)/2\)
and \( T \) is the distribution of \( dQ_2/d(Q_1 + Q_2) \) w.r.t \((Q_1 + Q_2)/2\).

(It follows directly from the testing criterion for comparison
that it suffices, in order to verify \((\epsilon_1, \epsilon_2)\) deficiency, to
consider functions \( \varphi \) of the form: \( x \rightarrow |x - \theta| \) where
\( \theta \in ]0,1[ \).)
Introduction.

Consider two dicholomies $\mathcal{G}$ and $\mathcal{F}$. The main result in this paper concerns the asymptotic behaviour of $\Delta(\mathcal{G}^n, \mathcal{F}^n)$ as $n \to \infty$. Here $\Delta$ denotes LeCam's distance for experiments and $n$-th power indicates $n$ independent replicates. The results of our investigations are given in sections 3 and 6. Sections 2, 4 and 5 are of an expository nature and is included for the sake of completeness as well as for rephrasing in terms of experiments.

It turns out that the theory of asymptotic behaviour of dichotomies is closely related to the theory of probabilities of very large deviations. Some background material on very large deviations is given in section 2. Two asymptotic expansions is given here. The first is based on the Berry-Esseen inequality and the latter on the Edgeworth expansion. We refer to volume II of Feller's book and Ibragimov and Linnik: "Independent and stationary sequences of random variables" for references on these results. Using the last expansion we derive Chernoff's theorem on the exponential rate of convergence of probabilities of very large deviations.

In section 3 we derive various expressions for deficiencies between dichotomies. The sublinear function criterion is extended to a criterion involving arbitrary convex (concave) functions. An immediate application is to the kernels of the Hellinger transformations. As a result we get a lower bound for exponential convergence of deficiencies. The standard probability measures are, essentially, the distribution of the posterior distribution under the uniform prior. We give formula's for deficiencies in terms of the cumulative distribu-
tion functions of these measures. The rest of this section is a collection of simple facts on laws of likelihood ratios, on probabilities of errors, on the boundary values of the Hellinger transforms, on representations and on expressions for deficiencies in terms of the integrated cumulative distribution functions of likelihood ratios.

The integrated cumulative distribution functions of likelihood ratios are closely related to Baye's risks for testing or, equivalently, to the maximum probability of guessing the true distribution. Asymptotique Baye's risk is the subject of section 4. The results given here (as well as in the next section) are well known. In particular we prove Chernoff's result on the rate of exponential convergence. This result states, essentially, that the $n$-th root of the Baye's risk converges to the minimum, $c(\mathcal{G})$, of the Hellinger transform of $\mathcal{G}$. Asymptotic expansions, analogous to those in section 2 are given.

Section 5 does the same thing for probabilities of errors. In particular we give Joshi's theorem on the convergence of the $n$-th root and an expansion due to Effron.

It is not too surprising that the two idempotent types of dichotomies play a particular role in this investigation. It is shown in section 6 that the distances to these types, i.e. the maximum informative type and the minimum informative type, satisfies two simple inequalities. Furthermore the distance to a minimum informative experiment is closely related to the maximal probability of guessing the true distribution for the uniform prior. Combining this with Chernoff's result described in section 3 we find finally that
\[ \sqrt[n]{\Delta(G^n, F^n)} \to \max(c(G), c(F)) \] as

\[ n \to \infty \] provided \( G \) and \( F \) are not equivalent. This shows the futility of trying to approximate an experiment \( G \) (not necessarily a dichotomy) by another \( F \), having more desirable properties, such that \[ \sqrt[n]{\Delta(G^n, F^n)} \to 0. \]
A symptotic behaviour of powers of dichotomies.

PD2. Some facts on very large deviations.


Consider independent and identically distributed random variables $X_1, X_2, \ldots$. Let $a$ be any real number. Then:

$$\lim_{n \to \infty} \sqrt[n]{P(X_n \geq a)} = \inf_{t \geq 0} E t(X_j - a)$$

where

$$X_n = \frac{1}{n}(X_1 + \ldots + X_n)$$

"\sqrt{}" may be replaced by "\inf{}" provided $P(X_j > a) > 0$ or $P(X_j = a) = 0$.

Furthermore it should be noted that - by the generalized Tschebyscheff inequality:

$$\sqrt[n]{P(X_n \geq a)} \leq \inf_{t \geq 0} E t(X_j - a)$$

for all $n$.

Proof of the last statement:

Let $t > 0$. Then:

$$P(X_1 + \ldots + X_n \geq na) = P(\sum_{j=1}^{n} (X_j - a) \geq 0)$$

$$= P(e^{t \sum_{j=1}^{n} (X_j - a)} \geq 1) \leq E e^{t \sum_{j=1}^{n} (X_j - a)}$$

$$= E \prod_{j=1}^{n} e^{t(X_j - a)} = [E e^{t(X_j - a)}]^{n}$$
Before proving the first statement let us consider the particular case where $E X_j = 0$, $a = \epsilon > 0$ and where there is a $t_0 \in ]0, \infty[$ so that

1) $E e^{tX_j} < \infty$ for some $t > t_0$ 

and

2) $E e^{t_0(X_j - \epsilon)} \leq E e^{t(X_j - \epsilon)}$ ; $t \in ]0, \infty[$

Then:

$$0 = \left[ \frac{d}{dt} E e^{t(X - \epsilon)} \right]_{t = t_0} = E(X - \epsilon)e^{t_0(X - \epsilon)}$$

Let $F$ denote the probability distribution of $X$ and let $H$ denote the probability distribution on $]-\infty, \infty[$ whose density w.r.t. $F$ is:

$$\gamma^{-1}e^{t_0(X - \epsilon)} ; x \in ]-\infty, \infty[$$

where $\gamma = E e^{t_0(X - \epsilon)} = \inf_t E e^{t(X - \epsilon)}$

Then

$$P(X \geq \epsilon) = \int_{[n, \infty[} (x_1 + \ldots + x_n) f_H^n(d(x_1, \ldots, x_n)) =$$

$$= \gamma^n \int_{[n, \infty[} (x_1 + \ldots + x_n) e^{-t_0[(x_1 - \epsilon) + \ldots + (x_n - \epsilon)]} f_H^n(d(x_1, \ldots, x_n)) =$$

$$= \gamma^n \int_0^\infty e^{-t_0} \gamma H_{-\epsilon}^n(dy) \text{ where } H_{-\epsilon} \text{ is the } (-\epsilon) \text{ translate of } H.$$ 

By partial integration:

$$P(X \geq \epsilon) = \gamma^n \int_0^\infty H_{-\epsilon}^n[0, \frac{1}{t_0} \log \frac{1}{z}]dz.$$
Furthermore:
\[ \int e^{TX}H(dx) = L(t_o + \eta)/L(t_o) \quad \text{where} \quad L(t) = e^{tX} \]

It follows that $H$ has finite moments of any order. In particular:

\[ \epsilon = \int xH(dx) = L'(t_o)/L(t_o) \]

and

\[ \int x^2H(dx) = L''(t_o)/L(t_o) \]

Let us write:

\[ \tau(\epsilon) = \sqrt{L(t_o)^{-1}L''(t_o) - \epsilon^2} \]

and

\[ \rho(\epsilon) = \sqrt{\int |x|^3H^{-\epsilon}(dx)} = \sqrt{L(t_o)^{-1}\int |x-\epsilon|^3e^{t_0X}P(dx)} \]

Clearly $\tau > 0$. By Zolotarev's sharpening [ ] of Berry-Essen's inequality we get:

\[ H_H^*[0, \frac{1}{t_o}\frac{\log}{\log Z}] = \Phi[0, \frac{1}{t_o}\frac{1}{t_0\sqrt{n}}] + \frac{2\rho^3}{\tau^3\sqrt{n}} \]

where $\Phi$ is the standard normal distribution and $|\theta| \leq 0.82$.

Hence

\[ P(\bar{X}_n \geq \epsilon) = \gamma^{n}\left\{ \int_0^{\epsilon} \Phi[0, \frac{1}{t_o}\frac{1}{t_0\sqrt{n}}]dz + \frac{2\rho^3}{\tau^3\sqrt{n}} \right\} \]

\[ = \gamma^{n}\left\{ \frac{1 - \Phi(t_o\sqrt{n})}{\sqrt{2\pi}} + \frac{2\rho^3}{\tau^3\sqrt{n}} \right\} \]
By the inequalities:
\[
\frac{1}{x} - \frac{1}{x^3} \leq \frac{1-x}{x^3(x)} \leq \frac{1}{x} ; \ x > 0
\]
we get:
\[
P(\bar{X}_n \geq \varepsilon) = \gamma^n \frac{1}{\tau \sqrt{n}} \int \frac{1}{\sqrt{2\pi}} \left(1 - \frac{\varepsilon}{t} + \frac{2\varepsilon^2}{t^2} \right)
\]

Let \( K \) be any non-lattice distribution such that
\[
\int xK(dx) = 0
\]
and
\[
\int |x|^3K(dx) < \infty
\]
Let \( t > 0 \) be a constant and put \( \tau = \sqrt{\int x^2K(dx)} \) and \( \kappa = \frac{3}{\int x^3K(dx)} \). Then
\[
\int \limits_0^1 k^{n^*}[0, \frac{1}{t} \log \frac{1}{y}]dy = \int \limits_0^\infty e^{-t\gamma k^{n^*}(dy)} = \frac{1}{\sqrt{2\pi n \tau}} (1 + o(1)) \cdot
\]

**Proof:**

We have seen that:
\[
\int \limits_0^1 k^{n^*}[0, \frac{1}{t} \log \frac{1}{y}]dy = \frac{1}{\sqrt{2\pi n}} \frac{1-x}{\tilde{g}'(tn)} = \frac{1}{\sqrt{2\pi n}} (1 + o(1)) \cdot
\]
In the general case we may use Edgeworth expansion of \( k^{n^*} \) as follows (Feller vol II):
\[
k^{n^*}(\sqrt{n} x) = \tilde{g}(x) + \frac{\kappa^3}{6 \sqrt{n}} (1-x^2) \tilde{g}'(x) + o(\frac{1}{\sqrt{n}})
\]
uniformly in \( x \in \mathbb{R} \).
Hence:

\[ \sqrt{n} \int_{0}^{1} H^{*}(0, \frac{1}{t} \log_{2} z) dz \]

\[ = \sqrt{n} \int_{0}^{1} \frac{1}{t} \log_{2} z dz \]

\[ + \frac{x^{3}}{6 \pi^{3} \sqrt{n}} \int_{0}^{1} \left[ 1 - \left( \frac{1}{t} \log_{2} z \right)^{2} \right] \frac{1}{t} \log_{2} z - \mathcal{E}(0) \] \(dz + o\left( \frac{1}{\sqrt{n}} \right) \). 

Writing \( b = t \sqrt{n} \tau \) and substituting \( z = e^{-by} \) in the integral, the middle term becomes:

\[ \int_{0}^{x} [(1 - y^{2}) \mathcal{E}(y) - \mathcal{E}(0)] e^{-by} dy = E[(1 - \frac{y^{2}}{b^{2}}) \mathcal{E}(y) - \mathcal{E}(0)] \]

where \( Y \) has density \( y \rightarrow e^{-y} \) on \([0, \infty[\). It follows that this integral \( \rightarrow 0 \) as \( b \rightarrow \infty \), and this completes the proof of statement (iii).

Applying this to \( P(\bar{X}_{n} \geq \epsilon) \) we get - provided \( F = \mathcal{L}(X_{i}) \) is not a lattice distribution:

\[ P(\bar{X}_{n} \geq \epsilon) = \frac{1}{\sqrt{2\pi n}} \int_{0}^{1} e^{-z^{2}/2} \frac{1}{t \log_{2} z} dz = \frac{1}{\sqrt{2\pi n}} \gamma^{*}(0) + o(1) \]

In particular:

\[ \lim_{n \rightarrow \infty} \sqrt{n} P(\bar{X}_{n} \geq \epsilon) = \gamma \]

We will repeatedly use the following simple minimax theorem (Chernoff).
Minimax theorem. Let $\chi$ be a compact topological space and $f_1 \leq f_2 \leq \cdots$ an infinite sequence of continuous extended real valued functions on $\chi$. Then

$$\inf \sup_{x} f_n(x) = \sup \inf_{x} f_n(x)$$

Note that it is not required that $\sup f_n$ is continuous. When we in the following apply this result to functions on $[0,\infty[$, then it is assumed that the functions are extended by continuity to $[0,\infty]$.

Consider now bounded random variables $X_1, X_2, \ldots$ and a constant $a$. We will demonstrate that

$$\inf_{n} \mathbb{P}(X_n \geq a) \to \inf_{t>0} \mathbb{E} e^t(X_j-a) \quad \text{We may - since this is trivial when } X_j \leq a - \text{assume that } \mathbb{P}(X_j > a) > 0$$

Then the variables $X_1-\mathbb{E}X_1, X_2-\mathbb{E}X_2, \ldots$ and the constant $\varepsilon = a-\mathbb{E}X_j$ satisfies 1) and 2) on page 2 provided $\mathbb{E}X_j < a$. Hence:

$$\liminf_{n} \sqrt{n} \mathbb{P}(X_n \geq a) = \liminf_{n} \sqrt{n} \mathbb{P}(X_n-\mathbb{E}X_j \geq \varepsilon)$$

$$= \inf_{t>0} \mathbb{E} e^{t(X_j-\mathbb{E}X_j-a)} = \inf_{t>0} \mathbb{E} e^{t(X_j-a)} \quad \text{provided } \mathbb{E}X_j < a$$

Suppose next that $\mathbb{E}X_j \geq a$. Choose $\delta > 0$ so that $\mathbb{P}(X_j = \mathbb{E}X_j+\delta) = 0$. Then

$$\liminf_{n} \sqrt{n} \mathbb{P}(X_n \geq a) \geq \liminf_{n} \sqrt{n} \mathbb{P}(X_n \geq \mathbb{E}X_j+\delta)$$

$$= \inf_{t>0} \mathbb{E} e^{t(X_j-\mathbb{E}X_j-\delta)} = \inf_{t>0} \mathbb{E} e^{t(X_j-\mathbb{E}X_j)} = 1 = \lim_{t>0} \mathbb{E} e^{t(X_j-a)}$$
Provided $P(X_j > E_Xj) > 0$. If $X_j = E_Xj$ then, trivially
\[ \inf_{t \geq 0} t(X_j - a) = 1 = \inf_{t \geq 0} E[e^{t(X_j - a)}]. \]

The general case is obtained by truncation as follows: Let $X_1, X_2, \ldots$ be independent and identically distributed random variables and $a$ a constant. Choose a number $w > a$ so that $P(|X_j| < w) > 0$. Put $A_w = \cap_{j=1}^{\infty} [|X_j| < w]$. Then
\[ P(\bar{X}_n \geq a) \geq P(|X_j| < w) \inf_{t \geq 0} E[e^{t(X_j - a)} | A_w]. \]

Hence:
\[ \liminf_{n} \frac{1}{n} \inf_{t \geq 0} \int_{|X_j| < w} e^{t(X_j - a)} \, dP \geq \liminf_{n} \frac{1}{n} \inf_{t \geq 0} E[e^{t(X_j - a)} | A_w] \]
\[ = \inf_{t \geq 0} \liminf_{n} \frac{1}{n} \inf_{t \geq 0} E[e^{t(X_j - a)}] \text{ as } w \uparrow \infty. \]

Suppose finally that $P(X_j > a) > 0$ or $P(X_j = a) = 0$. Choose $\delta > 0$ so that $P(X_j = a + \delta) = 0$. Then:
\[ \liminf_{n} \frac{1}{n} \inf_{t \geq 0} E[e^{t(X_j - a)}] \geq \liminf_{n} \frac{1}{n} \inf_{t \geq 0} E[e^{t(X_j - a - \delta)}] \]
\[ = \inf_{t \geq 0} E[e^{t(X_j - a - \delta)}] \uparrow \inf_{t \geq 0} E[e^{t(X_j - a)}]. \]

Altogether we have proved the:

The root theorem for large deviations: (Chernoff)

Let $X_1, X_2, \ldots$ be independent and identically distributed real valued random variables and let $a$ be a real number. Then
\[
\lim_{n \to \infty} \frac{\sqrt{n} \mathbb{P}(X_n \geq a)}{t(X_j-a)} = \inf_{t \geq 0} Ee
\]

"\geq" may be replaced by "\gt" provided \( P(X_j > a) > 0 \) or \( P(X_j = a) = 0 \).

For all \( n \):
\[
\frac{\sqrt{n} \mathbb{P}(X_n \geq a)}{t(X_j-a)} \leq \inf_{t \geq 0} Ee
\]

Extension to extended real valued random variables: With a view towards applications to log likelihood ratios let us note that the theorem holds when \( -\infty \) or \( +\infty \), but not both, is admitted as a possible value of \( X_j \).

References for section 2:

   On deviations of the sample mean. AMS 21, 1015-1027.

   The probability in the extreme tail of a convolution. AMS 30, 1113-1120.

   A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.
   AMS, 23, 493-507.

   Sur un nouveau théorème-limite de la théorie des probabilités.
   Act Sci et Ind 736 (1938).

   On the Berry-Esseen theorem.


Comparison of pseudo dichotomies in terms of functionals on the set of convex functions on $[0,1]$. (References: LeCam, Torgersen).

The purpose of this section is to provide a comparison criterion for pseudo dichotomies in terms of convex functions on $[0,1]$.

Let $S$ be a probability measure on $[0,1]$ such that $\int xS(dx) = 1/2$. Then $S$ represents the dichotomy $(S_1, S_2)$ where $\frac{dS_1}{dS} = 2(1-x)$ and $\frac{dS_2}{dS} = 2x$. There is an obvious 1-1 correspondence between standard probability measures (experiments) on $\{ (x_1, x_2) : x_1 \geq 0, x_2 \geq 0 \text{ and } x_1 + x_2 = 1 \}$ and probability measures $S$ (experiments $(S_1, S_2)$) on $[0,1]$.

Here is our comparison criterion in terms of convex functions:

**Theorem** Let $S$ and $T$ be probability measures on $[0,1]$ such that $\int xS(dx) = \int xT(dx) = 1/2$. Define probability measures $S_1, S_2, T_1$ and $T_2$ by:

$$\frac{dS_1}{dS} = \frac{dT_1}{dT} = 2(1-x)$$

$$\frac{dS_2}{dS} = \frac{dT_2}{dT} = 2x$$

Let $\epsilon_1, \epsilon_2 \geq 0$ be constants such that $(S_1, S_2)$ is $(\epsilon_1, \epsilon_2)$ deficient w.r.t. $(T_1, T_2)$. Let $\varphi$ be a convex function on $[0,1]$. Then - writing $0 \cdot \infty = 0$ - we have:

$$\epsilon_1 [\varphi'(1) - (\varphi(1) - \varphi(0))] + \epsilon_2 [\varphi(1) - \varphi(0)] - \varphi'(0) \geq \int \varphi d(T - S).$$
Conversely: if this inequality holds for all functions \( \varphi \) of the form

\[
\varphi(x) = |x-\theta|
\]

where \( \theta \in ]0,1[ \), then \((S_1,S_2)\) is \((\varepsilon_1,\varepsilon_2)\) deficient w.r.t. \((T_1,T_2)\).

\[ \text{Proof:} \] Suppose \((S_1,S_2)\) is \((\varepsilon_1,\varepsilon_2)\) deficient w.r.t. \((T_1,T_2)\) and let \( \varphi \) be convex on \([0,1]\). Define convex functions \( \varphi_n ; \ n = 1,2,\ldots \) by requiring that:

\[
\varphi_n(k/2^n) = \varphi(k/2^n) ; \ k = 0,1,\ldots, 2^n
\]

and that \( \varphi_n \) is linear on each interval \([i-1, i] 2^n \); \( i = 1,\ldots, 2^n \). Then \( \varphi_n \downarrow \varphi \), \( \varphi_n'(0) \downarrow \varphi'(0) \) and \( \varphi_n'(1) \uparrow \varphi'(1) \). It follows that we may - without loss of generality - assume that the graph of \( \varphi \) consists of a finite number of line segments. Hence we may write

\[
\varphi(x) = \max\{a_i+b_ix \ ; \ i = 1,\ldots,k\}
\]

where

\[
a_1 > a_2 > \ldots > a_k
\]

and

\[
b_1 < b_2 < \ldots < b_k .
\]

It follows that:

\[
\varphi(x) = a_1+b_1x + \sum_{i=1}^{k} (a_{i+1}+b_{i+1}x-a_i-b_i x)^+
\]
It suffices therefore - since the inequality is trivial when \( \varphi \) is linear - to prove it for functions \( \varphi \) of the form
\[
\varphi(x) = |x-\theta|; \quad \theta \in \mathbb{R}
\]. The inequality may then be written:

\[
||s_{1}-(1-\theta)s_{2}|| \geq ||t_{1}-(1-\theta)t_{2}|| - \varepsilon_{1}\theta - \varepsilon_{2}(1-\theta),
\]

and this follows from the testing criterion for comparison of dichotomies.

Suppose finally that (§) holds for \( \theta \in \mathbb{R} \). We must show that

\[
||a_{1}s_{1}+a_{2}s_{2}|| \geq ||a_{1}t_{1}+a_{2}t_{2}|| - \varepsilon_{1}|a_{1}| - \varepsilon_{2}|a_{2}|
\]

for any pair \((a_{1},a_{2})\) of constants. This is trivial when \( a_{1} \) and \( a_{2} \) have the same sign. Hence we may assume that \( a_{1} > 0 \) and \( a_{2} < 0 \) and then the inequality follows by inserting

\[
\theta = \frac{a_{1}}{a_{1}-a_{2}} \quad \text{in (§).}
\]

Let us next generalize the theorem to \( \delta_{1} \) equivalent and non-negative pseudo dichotomies.

Let \( \mathcal{C} = ((x,y), (\mu_{1},\mu_{2})) \) be a pseudo dichotomy such that \( \mu_{1} \geq 0 \) and \( \mu_{2} \geq 0 \). We may then define the standard measure of \( \mathcal{C} \) on \([0,1]\) as

\[
S \overset{\text{def}}{=} (\mu_{1}+\mu_{2})[d \mu_{2}/d(\mu_{1}+\mu_{2})]^{-1}
\]

Then \( \mathcal{C} \) is equivalent with the pseudo dichotomy \((s_{1},s_{2})\) where

\[
\frac{dS}{dS_{x}} = 1-x
\]

and
Clearly $S$ may be any finite measure on $[0,1]$. We formulate our theorem in terms of these measures:

**Theorem** Let $S$ and $T$ be finite measures on $[0,1]$ such that:

$$S[0,1] = T[0,1]$$
and

$$\int xS(dx) = \int xT(dx)$$

Define measures $S_1$, $S_2$, $T_1$ and $T_2$ by:

$$\left[ \frac{dS_1}{dx} \right]_x = 1 - x$$

$$\left[ \frac{dS_2}{dx} \right]_x = x$$

Let $\epsilon_1, \epsilon_2 \geq 0$ be constants such that $(S_1, S_2)$ is $(\epsilon_1, \epsilon_2)$ deficient w.r.t. $(T_1, T_2)$ and let $\varphi$ be convex on $[0,1]$. Then:

$$\epsilon_1[\varphi'(1) - (\varphi(1) - \varphi(0))] + \epsilon_2[(\varphi(1) - \varphi(0)) - \varphi'(0)] \geq 2\int \varphi d(T-S)$$

Conversely: if this inequality holds for all functions $\varphi$ of the form

$$\varphi(x) = |x-\theta|$$

where $\theta \in ]0,1[ , \text{ then } (S_1, S_2) \text{ is } (\epsilon_1, \epsilon_2) \text{ deficient w.r.t. } (T_1, T_2).$

**Proof:** Very similar to the proof of the previous theorem.
Example

Put, for each \( t \in ]0,1[ \), \( h_t(x) = (1-x)^{1-t}x^t \), \( x \in [0,1] \).

Then
\[
h_t'(x)/h_t(x) = \frac{t-x}{x(1-x)} \quad \text{and} \quad h_t''(x)/h_t(x) = -\frac{t(1-t)}{x^2(1-x)^2}
\]
when \( x \in ]0,1[ \). It follows that \(-h_t\) is convex. Direct application of the inequality to these functions yields, however, nothing since the left hand side = \( \infty \) when \( \varepsilon_1 > 0 \) or \( \varepsilon_2 > 0 \).

Let us approximate \( h_t \) by functions which behaves better at 0 and 1. Let \( 0 < a < b < 1 \) and put:
\[
h_{t,a,b}(x) = h_t(x) \quad \text{when} \quad x \in [a,b], \quad h_{t,a,b}(a) = h_t(a), \quad h_{t,a,b}(b) = h_t(b).
\]
Extend \( h_{t,a,b} \) to all of \([0,1]\) by requiring that it is linear on \([0,a]\) and \([b,1]\).

then \( h_{t,a,b}'(0) = (\frac{1-a}{a})^{1-t} \)

and \( h_{t,a,b}'(1) = -(\frac{b}{1-b})^t \)

It follows that:
\[
\frac{1}{4}[(\frac{1-a}{a})^{1-t}+(\frac{b}{1-b})^t] \delta(S,T) \geq (S-T)(h_{t,a,b})
\]
\[
= (S-T)(h_t)-S(h_t-h_{t,a,b})+T(h_t-h_{t,a,b})
\]
\[
\geq (S-T)(h_t)-2[h_t(a)v h_t(b)] \geq (S-T)(h_t)-2[a^t v (1-b)^{1-t}].
\]

Hence we have proved:
\[
\frac{1}{4}[(\frac{1-a}{a})^{1-t}+(\frac{b}{1-b})^t] \delta(S,T) \geq \frac{1}{4}I_{\alpha}(t)-I_{\alpha}(t)-2[a^t v (1-b)^{1-t}].
\]
If, in particular, \( a+b = 1 \) then we get:

\[
\frac{1}{2} \left[ \frac{(1-a)^{1-t}}{a} + \frac{(1-a)^t}{a} \right] \delta(g, \tau) \geq L_g(t) - L_\tau(t) - 2a^{t(1-t)}.
\]

Consider now the case of product experiments \( g^n \) and \( \tau^n \).

Let \( a \in \mathbb{R} \) and define \( a_n; n = 1, 2, \ldots \) by \( \frac{1-an}{a_n} = (\frac{1-a}{a})^n \).

Then

\[
a_n = \frac{a^n}{a_n + (1-a)^n} \text{ so that } n \sqrt{a_n} \to \frac{a}{1-a}.
\]

Suppose \( L_g(t) > L_\tau(t) \) and that \( \left[ \frac{a}{1-a} \right]^{t(1-t)} < L_g(t) \).

Then, for \( n \) sufficiently large, \( a_n < (\frac{a}{1-a} + \eta)^n \) where \( \eta > 0 \) satisfies \( \left[ \frac{a}{1-a} + \eta \right]^{t(1-t)} < L(t) \).

Applying the inequality above to \( g^n \) and \( \tau^n \) we get:

\[
2\left[ (1-a)^{n(1-t)} + \frac{(1-a)^{nt}}{a} \right] \delta(g^n, \tau^n) \geq L_g(t)^n - L_\tau(t)^n(\frac{a}{1-a} + \eta)^n(t^{n(1-t)})
\]

Hence

\[
\left[ (1-a)^{1-t} \right] \liminf_n \delta(g^n, \tau^n) \geq L_g(t) \frac{1}{\delta(g^n, \tau^n)} \geq L_g(t).
\]

It follows that:

\[
\liminf_n \delta(g^n, \tau^n) \geq L_g(t)^{(1-t)\lambda t} > 0.
\]

In particular:

\[
\liminf_n \Delta(g^n, \tau^n) > 0 \text{ when } \Delta(g, \tau) > 0.
\]
More precisely:

$$\liminf_{n} \Delta(\mathcal{E}^{n}, \mathcal{F}^{n}) \geq L_{\mathcal{E}}(t) \frac{1}{t \wedge (1-t)} v L_{\mathcal{F}}(t) \frac{1}{t \wedge (1-t)}$$

provided $L_{\mathcal{E}}(t) \neq L_{\mathcal{F}}(t)$. The provision may, since $\{ t : L_{\mathcal{E}}(t) = L_{\mathcal{F}}(t) \}$ is enumerable when $\mathcal{E} \neq \mathcal{F}$, be dropped when $\mathcal{E} \neq \mathcal{F}$. It follows from the convexity of $\log L_{\mathcal{E}}$ and $\log L_{\mathcal{F}}$ that the optimal choice of $t$ is $t = \frac{1}{2}$ and this choice of $t$ yield:

$$\liminf_{n} \Delta(\mathcal{E}^{n}, \mathcal{F}^{n}) \geq L_{\mathcal{E}}(\frac{1}{2})^2 v L_{\mathcal{F}}(\frac{1}{2})^2 > 0 \text{ when } \mathcal{E} \neq \mathcal{F}.$$ 

It will be shown in section 6, using methods based on the theory of probabilities of very large deviations, that the $\liminf$ actually is a $\lim$ and that this limit is

$$\max(\inf_{t} L_{\mathcal{E}}(t), \inf_{t} L_{\mathcal{F}}(t)).$$

The arguments above are, nevertheless, interesting since they are based solely on a few facts on statistical experiments.

Note also, by the inequality for $\liminf \sqrt[n]{\Delta(\mathcal{E}^{n}, \mathcal{F}^{n})}$, that

$$L_{\mathcal{E}} \leq L_{\mathcal{F}}$$

provided $\liminf_{n} \sqrt[n]{\Delta(\mathcal{E}^{n}, \mathcal{F}^{n})} = 0$. 
The standard probability measures are, essentially, the distributions of the posterior distributions under the uniform prior. It is therefore of considerable interest to express and interpret the basic facts in terms of the standard probability measures. The next theorem expresses deficiencies in terms of the integrated cumulative probability distributions.

**Theorem** Let $S$ and $T$ be standard probability measures on $[0,1]$. Then $S$ (i.e. the experiment defined by $S$) is $(\epsilon_1, \epsilon_2)$ deficient w.r.t. $T$ if and only if:

$$\epsilon \epsilon_1 + (1-\epsilon) \epsilon_2 \geq 4 \left( \tilde{T}(\epsilon) - \tilde{S}(\epsilon) \right) ; \epsilon \in ]0,1[$$

where $\tilde{T}(\epsilon) = \int_0^\epsilon T(t) dt$ and $\tilde{S}(\epsilon) = \int_0^\epsilon S(t) dt$.

**Proof:** By partial integration:

$$\int |x-\epsilon|S(dx) = \frac{3}{2} - \epsilon + 2\tilde{S}(\epsilon)$$

for any standard probability measure $S$ on $[0,1]$. The theorem follows now from the comparison criterion above. □

**Remark:** Note that $\tilde{S}(\epsilon) = \frac{\epsilon}{2}$ or $= (\epsilon - \frac{1}{2})^+$ as $S$ is minimal or maximal. A statistical interpretation of the integral $\tilde{S}(\epsilon)$ may be obtained by observing that $\tilde{S}(\epsilon) = \frac{\epsilon}{2}[\epsilon - B(1-\epsilon)]$ where $B(1-\epsilon)$ is the minimum Bayes' risk for the problem of testing $S_1$ against $S_2$ for 0-1 loss and prior $(\epsilon, 1-\epsilon)$. 
Corollary

\[ \delta(S, T) = 4 \sup_{0 < \epsilon < 1} \left[ \overline{T}(\epsilon) - \overline{S}(\epsilon) \right] + \]

and

\[ \Delta(S, T) = 4 \sup_{0 < \epsilon < 1} \left| \overline{T}(\epsilon) - \overline{S}(\epsilon) \right| \]

Corollary

\[ \delta(S, T) = 0 \text{ if and only if } \overline{S} \geq \overline{T}. \]

More generally - consider finite non negative measures S and T on \([0, 1]\) so that

\[ \|S\| = \|T\| \]

and

\[ \int xS(dx) = \int xT(dx) \]

Define measures \(S_1, S_2, T_1\) and \(T_2\) by

\[ \frac{dS_1}{dS(x)} = 1 - x = \frac{dT_1}{dT(x)} \]

\[ \frac{dS_2}{dS(x)} = x = \frac{dT_2}{dT(x)} \]

Then the pseudodichotomy \((S_1, S_2)\) is \((\epsilon_1, \epsilon_2)\) deficient w.r.t. \((T_1, T_2)\) if and only if:

\[ \epsilon_1 + (1-\epsilon)\epsilon_2 \geq 2(\overline{T}(\epsilon) - \overline{S}(\epsilon)) ; \ \epsilon \in ]0, 1[ \]

where \(\overline{T}(\epsilon) = \int_0^\epsilon T[0,t]dt\) and \(\overline{S}(\epsilon) = \int_0^\epsilon S[0,t]dt\).
Let us next summarize a few facts on comparison in terms of laws of likelihood ratios.

Let \( \mathcal{L} = (x, P, Q) \) be any dichotomy. The standard probability measure of \( \mathcal{L} \) is \( S = \int \frac{dQ}{d(P+Q)} \). This measure determines and is determined by \( K = \int \frac{dQ}{dP} \).

S integrals may be converted into K integrals and conversely by the formula's:

\[
\int h dK = 2 \int h(x) (1-x) S(dx) \quad h \in C[0, \infty] \quad x < 1
\]

and

\[
\int h dS = h(1) + \int [h(x) - h(1)] (1+x) K(dx)
\]

In particular:

\[
K(z) = 2 \int (1-x) S(dx)
\quad [0, \frac{z}{1+z}]
\]

K is also related to probabilities of errors of the first and the second kind through the formula:

\[
1 - \beta_\alpha = \int_0^1 K^{-1}(1-p) dp
\]

where \( \beta_\alpha \) is the power of the most powerful level \( \alpha \) test for testing \( P \) against \( Q \).

Let \( \varphi \) be continuous and convex on \([0, 1]\). Then:

\[
\int \varphi dS = \varphi(1) + \int \psi dK
\]

where

\[
\psi(x) = [\varphi(x) - \varphi(1)] (1+x)
\]

Then \( \psi \) is convex and continuous on \([0, \infty] \) and \( \psi(z)/z \to 0 \) as \( z \to \infty \). Conversely: Any continuous convex function \( \psi \) on \([0, \infty] \) such that \( \psi(z)/z \to 0 \) as \( z \to \infty \) may be written:
\[ \psi(x) = [\phi\left(\frac{x}{1+x}\right) - \phi(1)](1+x); \ x \geq 0 \]

where \( \phi \) is continuous and convex on \([0,1]\). \( \phi \) determines \( \psi \) and is determined by \( \psi \) up to an additive constant.

The Hellinger transform, \( L \), of \( \varphi \) may be expressed by \( K \) by:

\[ L(t) = \int x^+ K(dx); \ t \in ]0,1[ \]

Before proceeding let us note the following properties of the Hellinger transform:

- \( L(0^+) = 1 - K(0) = \) the total variation of the \( Q \) absolutely continuous part of \( P \).
- \( L(1-) = \int x K(dx) = \) the total variation of the \( P \) absolutely continuous part of \( Q \).
- \( L(r)(t) = \int (\log x)^r x^+ K(dx); \ 0 < t < 1 \)
- \( \lim_{t \to 0^+} L'(t) = \lim_{h \to 0} \frac{L(h) - L(0^+)}{h} = \mathbb{E}_P[\frac{dQ}{dP} \log \frac{dQ}{dP}]_{]0,\infty[^0} \)
- \( \lim_{t \to 1^-} L'(t) = \lim_{h \to 0} \frac{L(1-h) - L(1^-)}{-h} = \mathbb{E}_P[\log \frac{dQ}{dP}] \frac{dQ}{dP} \)

The "metric" is obtained from \( K \) through the formula:

\[ ||\lambda P - Q|| = \int |\lambda - x| K(dx) + 1 - \int x K(dx) = 1 - \lambda + 2 \tilde{K}(\lambda); \ \lambda > 0, \]

where \( \tilde{K}(\lambda) = \frac{\lambda}{0} \int K(x) dx; \ \lambda > 0. \)

A representation of \( \varphi \) may be obtained from \( K \) by taking \([0,\infty[^0\) as sample space and putting \( P = K \) and \( Q(B) = \int x K(dx) + [1 - \int x K(dx)] I_B(\infty) \) for each Borel set \( B \).

Clearly \( \int x K(dx) \leq 1 \) and it follows from this construction that \( K \) may be any probability measure on \([0,\infty[^0\) so that \( \int x K(dx) \leq 1 \). It follows that \( \tilde{K} \) is and may be any continuous convex function on \([0,\infty[^0\) such that \( (\lambda - 1)^+ \leq \tilde{K}(\lambda) \leq \lambda, \lambda > 0 \).
Consider now two dichotomies \( \mathcal{E} = (x, A, P_1, P_2) \) and 
\( \mathcal{F} = ( (y, B), Q_1, Q_2 ) \). Put \( K = \mathcal{L}_{P_1} (dP_2/dP_1) \) and 
\( H = \mathcal{L}_{Q_1} (dQ_2/dQ_1) \). Then \( \mathcal{E} \) is \((\varepsilon_1, \varepsilon_2)\) deficient w.r.t. \( \mathcal{F} \) if and only if 
\[
\varepsilon_1 \lambda + \varepsilon_2 \geq 2 (R(\lambda) - K(\lambda)) ; \lambda > 0
\]
where \( R(\lambda) = \int_{0}^{\lambda} K(x)dx \) and \( R(\lambda) = \int_{0}^{\lambda} H(x)dx \).

Hence:
\[
\delta(\mathcal{E}, \mathcal{F}) = 2 \sup \frac{R(\lambda) - K(\lambda)}{1 + \lambda}^+
\]
\[
\Delta(\mathcal{E}, \mathcal{F}) = 2 \sup \frac{|R(\lambda) - K(\lambda)|}{1 + \lambda}
\]

In particular
\[
\delta(\mathcal{E}, \mathcal{F}) = 0 \text{ if and only if } R \geq \bar{R}.
\]

References for section 3:


Consider the problem of testing \( P \) against \( Q \) in a dichotomy \( \xi = ((x, \mathcal{A}), P, Q) \). Any prior distribution may be identified with the probability \( \lambda \) it assigns to the second coordinate of the ordered pair \((P, Q)\). To avoid trivialities we will - unless otherwise stated - assume that \( \lambda \) is not degenerate i.e. that:

\[
0 < \lambda < 1
\]

Consider now the 0-1 loss function

<table>
<thead>
<tr>
<th>accept</th>
<th>reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td>0</td>
</tr>
<tr>
<td>( Q )</td>
<td>1</td>
</tr>
</tbody>
</table>

Then any test with rejection region \( \left[ \frac{dQ}{dP} > \frac{1-\lambda}{\lambda} \right] \) where \( \frac{dQ}{dP} \) is \( Q \)-maximal, minimizes the Baye's risk. Equivalently the guessing strategy: "Guess \( P \) or \( Q \) as \( \frac{dQ}{dP} \leq \frac{1-\lambda}{\lambda} \) or > \( \frac{1-\lambda}{\lambda} \)" maximizes the probability of being right.

The minimum Baye's risk, \( B_\lambda(\xi) \), and the maximum probability of guessing the right distribution, \( M_\lambda(\xi) \) are:

\[
B_\lambda(\xi) = \| (1-\lambda)P \land \lambda Q \|
n\]

and

\[
M_\lambda(\xi) = \| (1-\lambda)P \lor \lambda Q \|
\]

clearly

\[
B_\lambda(\xi) + M_\lambda(\xi) = 1.
\]
A simple upper bound for \( B_\lambda (\frac{C}{G}) \) is given by:

\[
B_\lambda (\frac{C}{G}) \leq \inf_{0 < t < 1} (1 - \lambda)^{1-t} \lambda^t I_G (t)
\]

where \( I_G (t) = \int dP^1 - t dQ^t ; t \in ]0, 1[ \)

is the Hellinger transform of \( \frac{C}{G} \).

It follows that:

\[
B_\lambda (\frac{C^n}{G}) \leq \inf_{0 < t < 1} (1 - \lambda)^{1-t} \lambda^t [I_G (t)]^n .
\]

Chernoff, 1952, proved - using Cramer's results on probabilities of large deviations - that:

\[
\lim_{n \to \infty} \frac{\sqrt{n}}{B_\lambda (\frac{C^n}{G})} = \inf_{0 < t < 1} I_G (t)
\]

By the inequality above it suffices to show that

\[
\liminf_{n \to \infty} \frac{n}{\sqrt{B_\lambda (\frac{C^n}{G})}} \geq \inf_{0 < t < t} I_G (t) .
\]

Put \( a = \log (\frac{1-\lambda}{\lambda}) \) and let \( Z_i = \log \frac{dQ}{dP} \) in the \( i \)-th trial. Consider first the case where there is a \( t_0 \in ]0, 1[ \) so that

\[
0 < I_G (t_0) \leq I_G (t) < 1 ; t \in ]0, 1[ .
\]

It is easily seen that \( I_G (t) = E e^{tZ} \) where \( Z \) may be any version of \( \log \frac{dQ}{dP} \). By differentiation:
0 = L'(t_0) = \int z e^{t_0 z} dP

The inequality \( L(t_0) < 1 \) implies that \( P(Z > -\infty) > 0 \). Denote by \( F \) the conditional distribution, under \( P \), of \( Z \) given that \( Z > -\infty \). Then

\[
L(t) = \eta \int e^{t z} F(dz)
\]

where \( \eta = P(Z > -\infty) \). Let \( G \) be the probability distribution on \( \mathbb{R} \) whose density w.r.t. \( F \) is

\[
\left[ \int e^{t_0 z} F(dz) \right]^{-1} e^{t_0 z} = L(t_0)^{-1} \eta e^{t_0 z}; z \in \mathbb{R}.
\]

It is then easily seen that the Baye's risk is:

\[
B_\lambda(\mathcal{C}) = \lambda \eta \int e^{z \lambda} \mathcal{C} F(dz).
\]

Hence:

\[
B_\lambda(\mathcal{C}^n) = \| (1-\lambda) F \wedge \lambda G^n \| = \lambda \eta \int e^{(z_1 + \ldots + z_n) \lambda - t_0 (z_1 + \ldots + z_n)} F^n(dz)
\]

\[
= \lambda L(t_0)^n \int e^{(z_1 + \ldots + z_n) \lambda - t_0 (z_1 + \ldots + z_n)} G^n(dx)
\]

\[
= \lambda L(t_0)^n \int_0 e^{(z_1 + \ldots + z_n) \lambda - t_0 (z_1 + \ldots + z_n)} G^n(dx)
\]

A few moments of \( G \) are:

\[
\int z G(dz) = 0
\]
\[ I = \int_{-a(1-t_0)}^{\infty} \left[ (1-\left(\frac{a-\log y}{s}\right)^2) \frac{a-\log y}{s} \right] dy \]

and where \( s = \tau/n \).

Substituting \( z = \log \frac{1}{y} \) in the integral we get:

\[ I = \int_{-a(1-t_0)}^{\infty} \left[ (1-\left(\frac{z+a}{s}\right)^2) \frac{z+a}{s} \right] e^{-z} \, dz \]

\[ \rightarrow \int_{-a(1-t_0)}^{\infty} [z'(0)-\lambda'(0)] e^{-z} \, dz = 0 \text{ as } s \to \infty. \]

Hence (Efron and Truax [ ])

\[ B_\lambda (C^n) = I(t_0)^n \left(1-\frac{1-t_0}{1-t}\right) \frac{1}{\sqrt{2\pi t^2}} (1+o(1)) \text{ as } n \to \infty \]

provided \( \frac{dQ}{dP} \) is not a lattice distribution.

The functions \( \lambda \rightarrow B_\lambda (C^n) \) and \( \lambda \rightarrow (1-\lambda)^{1-t_0} \) are both concave and continuous on \([0,1]\). It follows that the convergence is uniform on \([0,1]\).

Consider now the general case. Let \( Z_j \) and \( V_j \) be, respectively versions of \( \log \frac{dP}{dQ} \) and \( \log \frac{dQ}{dP} \) in the \( j \)-th trial. Then:

\[ B_\lambda (C^n) = (1-\lambda)P^n(\zeta_n \geq \frac{a}{n}) + \lambda Q^n(\zeta_n \geq - \frac{a}{n}) \]
where $Z_n = \frac{1}{n} \sum_{j=1}^{n} Z_j$ and $V_n = \frac{1}{n} \sum_{j=1}^{n} V_j$

By the theory of large deviations:

$$\frac{n}{\sqrt{P(Z_n > \frac{a}{n})}} \to \inf_{t \geq 0} E_{P} e^{tZ}$$

provided $P(Z > 0) > 0$ or $a \leq 0$ and

$$\frac{n}{\sqrt{Q(V_n > -\frac{a}{n})}} \to \inf_{t \geq 0} E_{Q} e^{tV}$$

provided $Q(V > 0) > 0$ or $a \geq 0$.

Consider now the functions $N(t) = E_{P} e^{tZ}$ and $M(t) = E_{Q} e^{tV}$.

Then $M(1-t) = N(t) = L(t)$ when $0 < t < 1$. Furthermore:

$N(M)$ is finite, continuous and convex in any interval $]0, t[$ such that $N(t) < \infty$ ($M(t) < \infty$).

$N(0+) = P(Z > -\infty)$,

$M(0+) = Q(V > -\infty)$,

$$\int Z dP = \lim_{Z \to \infty} N'(t) \leq \lim_{t \to 0} N'(t) = \int Z e^Z dP = -\int V dQ,$$

and

$$\int V dQ = \lim_{V \to \infty} M'(t) \leq \lim_{t \to 0} M'(t) = \int V e^V dQ = -\int Z dP$$
Suppose now that $\int e^Z dP \geq 0$ and $P(Z > 0) > 0$. Then

$$\lim_{n \to \infty} \inf_{t \geq 0} E e^{t Z} = \inf_{0 < t < 1} L(t).$$

By the first $n$-th root limit above:

$$\lim_{n \to \infty} \sqrt[n]{B_\lambda(\mathcal{G}^n)} = \inf_{0 < t < 1} L(t).$$

By a symmetric argument this holds also when $\int v e^V dQ \geq 0$ and $Q(V > 0) > 0$.

It remains to consider the case $Z \leq 0$ and $V \leq 0$. In this case $\mathcal{G}$ is equivalent with the experiment:

<table>
<thead>
<tr>
<th>$\mathcal{G}$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>1-p</td>
<td>p</td>
<td>0</td>
</tr>
<tr>
<td>$Q$</td>
<td>0</td>
<td>p</td>
<td>1-p</td>
</tr>
</tbody>
</table>

Hence $\mathcal{G}^n$ is equivalent with $\mathcal{G}^n_p$ so that

$$B_\lambda(\mathcal{G}^n) = p^n[\lambda(1 - \lambda)]$$

and

$$L_\mathcal{G}(t) = p; \ 0 < t < 1$$

Hence we have proved:

**Theorem (Chernoff)** $\sqrt[n]{B_\lambda(\mathcal{G}^n)} \to \inf_{0 < t < 1} L_\mathcal{G}(t)$ for any dichotomy $\mathcal{G}$.

**Remark.** It follows from the inequality $B_\lambda(\mathcal{G}^n) \leq \inf L_\mathcal{G}(t)^n$ and the concavity of the functions $B_\lambda(\mathcal{G}^n)$ that the convergence is uniform on compacts not containing 0 and 1.
Reference for section 4.

A measure of asymptotique efficiency for tests of a hypothesis based on the sum of observations.
AMS, 23, 493-507.

Deviation theory in exponential families.
AMS 39, 1402-1424.

Some inequalities of probability theory and their application to the refinement of A.M. Liapunov's theorem.
Doklady Akademii nauk. SSSR. 177 no. 3. 501-504.
Asymptotic behaviour of probabilities of errors of the first and the second kind. (References: Joshi, Kullback, Efron, Rao, Zolotarev.)

Let \( G = (\mathcal{X}, \mathcal{T}, P, Q) \) be a given dichotomy. Denote \( \beta_n(\alpha) \) the power of the most powerful level \( \alpha \) test for testing \( P^n \) against \( Q^n \). Then \( 1-\beta_n(\alpha) \) is the minimum probability of error of the second kind given that the probability of an error of the first kind is at most \( \alpha \). We shall in this section study the asymptotic behaviour of \( \beta_n(\alpha) \) as \( n \to \infty \). We will - since \( \beta_n(1) = 1 \) and \( 1-\beta_n(0) = (1-\beta(0))^n \) assume that \( 0 < \alpha < 1 \).

Important information on the asymptotic behaviour of \( 1-\beta_n(\alpha) \) may be obtained from:

**Theorem**

Suppose \( 0 < \alpha < \beta(\alpha) < 1 \) and that \( c \) is a constant so that:

\[
P\left(\frac{dP}{d\mathcal{T}} \geq c\right) \geq \alpha \geq P\left(\frac{dP}{d\mathcal{T}} > c\right)
\]

Then \( c > 0 \). Furthermore:

\[
\frac{1}{d}[1-\alpha) - P\left(\frac{dP}{d\mathcal{T}} \leq \frac{1}{d}\right)] \leq 1-\beta(\alpha) \leq c(1-\alpha) \quad \text{when} \quad cd > 1
\]

**Proof:** Let \( \frac{dQ}{d\mathcal{T}} \) denote a \( Q \)-maximal version. \( c \leq 0 \) imply \( \beta(\alpha) = 1 \). Hence \( 0 < c < \infty \). Choose \( \gamma \in [0,1] \) so that

\[
1-\alpha = P\left(\frac{dQ}{d\mathcal{T}} < c\right) + (1-\gamma)P\left(\frac{dQ}{d\mathcal{T}} = c\right)
\]
By the Neyman-Pearson lemma:

$$1 - \beta(\alpha) = Q(\frac{dQ}{dP} < c) + (1 - \gamma)Q(\frac{dQ}{dP} = c)$$

Hence:

$$1 - \beta(\alpha) = \int_{\frac{dQ}{dP} < c} dQ + (1 - \gamma)\int_{\frac{dQ}{dP} = c} c dP = c (1 - \alpha)$$

and

$$1 - \beta(\alpha) \leq Q\left(\frac{1}{d} < \frac{dQ}{dP} < c\right) + (1 - \gamma)Q\left(\frac{dQ}{dP} = c\right) = \frac{1}{d}(1 - \alpha - P(\frac{dQ}{dP} \leq \frac{1}{d})) .$$

Suppose now that $0 < \alpha < \beta_n(\alpha) < 1$ for $n = 1, 2, \ldots$.

To each $n = 1, 2, \ldots$ there is, by the theorem, a $c_n > 0$ so that

$$P^n(\frac{dQ^n}{dP^n} \geq c_n) \geq \alpha \geq P^n(\frac{dQ^n}{dP^n} > c_n) .$$

Let $Z$ denote a version of $\log\frac{dQ}{dP}$ and $Z_j$ a version of $\log\frac{dQ}{dP}$ from the $j$-th trial. Then:

$$P^n(\sum_{j=1}^{n} Z_j \geq \log c_n) \geq \alpha \geq P^n(\sum_{j=1}^{n} Z_j > \log c_n)$$

or

$$P^n(Z \geq \log \sqrt{n}c_n) \geq \alpha \geq P^n(Z > \log \sqrt{n}c_n) .$$
By the weak law of large numbers:

\[ \frac{1}{n} \sum_{i=1}^{n} Z_i \to EZ \text{ as } n \to \infty. \]

It follows that \( \log \sqrt[n]{c_n} \to E_Z \). We do not need to assume that \( Z \) is \( P \) integrable. By the theorem:

\[ 1 - \beta_n(\alpha) \leq c_n(1 - \alpha) \]

Hence

\[ \sqrt[n]{1 - \beta_n(\alpha)} \leq \sqrt[n]{c_n} \sqrt[n]{1 - \alpha} \]

so that

\[ \limsup_n \sqrt[n]{1 - \beta_n(\alpha)} \leq e^{E_Z} \]

Suppose next that \( E_Z > -\infty \). Choose \( \epsilon > 0 \). By the theorem:

\[ 1 - \beta_n(\alpha) \geq e^{n(E_Z - \epsilon)} \left[ 1 - P(Z_n \leq E_Z - \epsilon) \right] \]

when \( n \) is sufficiently large. Hence

\[ \liminf_n \sqrt[n]{1 - \beta_n(\alpha)} \geq e^{E_Z - \epsilon} \to e^{E_Z} \]

We have almost proved:

**Theorem (Joshi)**

\[ \lim_{n \to \infty} \sqrt[n]{1 - \beta_n(\alpha)} = e^{E \log \frac{dQ}{dP}} \]

provided \( 0 < \alpha < 1 \). 

-------------------
Remark 1

\[-E_P \log \frac{dQ}{dP} = -L'(o)\] provides a non negative, monotone and additive (w.r.t. products) measure of information.

Remark 2

The convergence is automatically uniform on intervals \([a_0, a_1]\) where \(0 < a_0 < a_1 < 1\).

Completion of the proof of the theorem: It remains to consider the case where \(\beta_n(a) = 1\) for some \(n\). In this case \(Q \not\succ P\) so that \(E_P \log \frac{dQ}{dP} = -\infty\).

We may get better estimates of \(\beta_n(a)\) provided we make assumptions on the moments of \(\log \frac{dQ}{dP}\) under \(P\). A simple consequence of the central limit theorem is:

**Theorem.**

Suppose \(E_P \log \frac{dQ}{dP} > -\infty\) and that \(0 < \text{Var}_P \log \frac{dQ}{dP} < \infty\).

Let \(0 < a < 1\). Then:

\[1 - \beta_n(a) = e^{[E_P \log \frac{dQ}{dP}]n+1(1-a)} \sqrt{\text{Var}_P \log \frac{dQ}{dP}} \sqrt{n} + o(\sqrt{n})\]

uniformly in \(a \in [a_0, a_1]\) provided \(0 < a_0 < a_1 < 1\).

**Proof:** Write \(Z = \log \frac{dQ}{dP}\), \(\xi = E_P Z\) and \(\tau = \text{Var}_P Z\). Let \(Z_j\) be the value of \(Z\) in the \(j\)-th trial. Our assumptions imply that there is, for each \(n\), a positive constant \(c_n\) so that:

\[P^n(\xi_n \geq \log \frac{n}{\sqrt{c_n}}) \geq c_n \geq P^n(\xi_n > \log \frac{n}{\sqrt{c_n}})\]
where $Z_n = \frac{1}{n} \sum_{j=1}^{n} Z_j$. Put $r_n = \log \sqrt{c_n}$. The inequalities may be written:

$$P^n\left(\frac{Z_n - \zeta}{\sqrt{n}} \geq \frac{r_n - \zeta}{\sqrt{n}}\right) \geq \alpha \geq P^n\left(\frac{Z_n - \zeta}{\sqrt{n}} < \frac{r_n - \zeta}{\sqrt{n}}\right).$$

By the central limit theorem:

$$\frac{r_n - \zeta}{\sqrt{n}} \rightarrow \Phi^{-1}(1-\alpha)$$

uniformly in $\alpha \in [\alpha_0, \alpha_1]$.

By the basic inequalities:

$$1 - \beta_n(\alpha) \leq c_n(1-\alpha)$$

or

$$\sqrt{n} \left[\frac{1}{n} \log(1 - \beta_n(\alpha)) - \zeta\right] \leq \frac{r_n - \zeta}{\sqrt{n}} + \log(1-\alpha)$$

Hence

$$\limsup_n \sqrt{n} \left[\frac{1}{n} \log(1 - \beta_n(\alpha)) - \zeta\right] \leq \tau \Phi^{-1}(1-\alpha); \text{ uniformly in } \alpha \in [\alpha_0, \alpha_1].$$

Put $\frac{1}{n} \log \frac{1}{d_n} = \zeta + \frac{\tau \Phi^{-1}(1-\alpha) - \eta}{\sqrt{n}}$ where $\eta > 0$ is a constant.

Then:

$$\frac{1}{n} \log c_n + \frac{1}{n} \log d_n = \frac{\phi(1) + \eta}{\sqrt{n}} > 0; \quad \alpha \in [\alpha_0, \alpha_1]$$
when \( n \) is sufficiently large. It follows that there is a \( N \) so that \( c_n d_n > 1 \) for all \( \alpha \in [\alpha_0, \alpha_1] \) provided \( n \geq N \).

By the basic inequalities again:

\[
\frac{1}{n} \log (1 - \beta_n(\alpha)) \geq \frac{1}{n} \log \frac{1}{d_n} \log [1 - \alpha - P_n(\frac{\zeta}{\sqrt{n}})]^+ \\
= \zeta + \frac{\tau^{-1}(1-\alpha)-\eta}{\sqrt{n}} + \frac{1}{n} \log [1 - \alpha - P_n(\frac{\zeta}{\sqrt{n}})]^+ \\
= \zeta + \frac{\tau^{-1}(1-\alpha)-\eta}{\sqrt{n}} + \frac{1}{n} \log [1 - \alpha - P_n(\frac{\zeta}{\sqrt{n}})]^+
\]

when \( n \geq N \). It follows that:

\[
\liminf_n \sqrt{n} \left[ \frac{1}{n} \log (1 - \beta_n(\alpha)) - \zeta \right] \geq \tau^{-1}(1-\alpha) - \eta, \text{ uniformly in } \alpha \in [\alpha_0, \alpha_1]. \eta \to 0 \text{ yield}
\]

\[
\liminf_n \sqrt{n} \left[ \frac{1}{n} \log (1 - \beta_n(\alpha)) - \zeta \right] \geq \tau^{-1}(1-\alpha), \text{ uniformly in } \alpha \in [\alpha_0, \alpha_1].
\]

Hence:

\[
\sqrt{n} \left[ \frac{1}{n} \log (1 - \beta_n(\alpha)) - \zeta \right] \to \tau^{-1}(1-\alpha), \text{ uniformly in } \alpha \in [\alpha_0, \alpha_1].
\]

Let \( \frac{dQ^n}{dP^n} ; n = 1, 2, \ldots \) denote \( Q^n \) - maximal versions. By the first part of the proof, the most powerful level \( \alpha \) test rejects or accepts as \( \frac{dQ^n}{dP^n} > c_n \) where

\[
\log c_n = \left[ E_P \log \frac{dQ}{dP} \right] + \tau^{-1}(1-\alpha) \sqrt{\text{Var}_P \log \frac{dQ}{dP}} \sqrt{n + o(\sqrt{n})}
\]

uniformly in \( \alpha \in [\alpha_0, \alpha_1] \), and where \( 0 < \alpha_0 < \alpha_1 < 1 \). In order to get better estimates of \( \beta_n(\alpha) \) we need better estimates of \( c_n \).
Suppose $E_P[\log \frac{dQ}{dP}]^2 < \infty$. Then Effron \[ \] established an asymptotic expression for $\beta_n(\alpha)$ which is considerably more efficient than the result above. We will derive Effron's expansion below. The main difference between our proof and the proof in \[ \] is the method of estimation of the crucial integral $I_n$ (to be defined below).

We keep the notations of the proof of the last theorem and we will assume that $\alpha \in [a_0, a_1]$ where $0 < a_0 < a_1 < 1$ are given numbers. All limits in the following arguments are, if not otherwise stated, uniform in this interval. Suppose first that $\mathcal{L}_P(Z)$ is not a lattice distribution.

Let $F_n$ be the distribution of $T_n \overset{\text{def}}{=} \frac{Z_n - \xi}{\tau \sqrt{n}}$ under $P_n$. Then

$$F_n(x) = \phi(x) + \frac{\kappa^3}{6\tau^3\sqrt{n}} R(x) + o(1) \sqrt{n},$$

where $R(x) = (1-x^2)\phi'(x)$ and $\kappa = \frac{3\mathcal{L}_P(Z - \xi)^3}{\sqrt{P_n}}$.

Writing $r_n = \log \frac{r_n}{\sqrt{c_n}}$ we get:

$$F_n(b_n) \leq 1 - \alpha \leq F_n(b_n +)$$

where $b_n = \frac{r_n - \xi}{\tau \sqrt{n}}$. Put $s_\alpha = \phi^{-1}(1-\alpha)$ and $\delta_n = b_n - s_\alpha$.

The expansion of $c_n$ above imply that $\delta_n = o(1)$.

The expansion of $F_n$ yield:

$$1 - \phi(s_\alpha + \delta_n) - \frac{\kappa^3}{6\tau^3\sqrt{n}} R(s_\alpha + \delta_n) = \alpha + o(1) \sqrt{n}$$

so that
\[ 1 - \alpha - \delta(s + \xi) - \frac{3\kappa^3}{6\tau^2 \sqrt{n}} R(s) = o(1) / \sqrt{n} \]

or

\[ - \delta_n \left[ \delta'(s) + o(1) \right] - \frac{3\kappa^3}{6\tau^2 \sqrt{n}} R(s) = o(1) / \sqrt{n} \]

Hence

\[ \delta_n = \frac{3\kappa^3}{6\tau^2 \sqrt{n}} (1 - s^2) + o(1) / \sqrt{n} \]

Solving w.r.t. \( c_n \) we get:

\[ \log c_n = \xi + s_n \sqrt{n} + \frac{3\kappa^3}{6\tau^2} (1 - s_n^2) + o(1) \]

Using this estimate of \( c_n \) we may get a better estimate of \( \beta_n(\alpha) \) as follows:

By Neyman-Pearson's lemma:

\[ Q_n(T_n < b_n) \leq 1 - \beta_n(\alpha) \leq Q_n(T_n \leq b_n) \]

Consider now \( Q_n(T_n(\leq) b_n) = F_n(b_n(+) \]

The formula \( \bar{z}_n = 6 + \tau \frac{T_n}{\sqrt{n}} \) yield:

\[ Q_n(T_n(\leq) b_n) = \int_{T_n(\leq) b_n} e^{n\xi + \sqrt{n} \tau} dF_n = \]

\[ = e^{n\xi + \sqrt{n} \tau b_n} \frac{1}{\sqrt{n} \tau} \int_{T_n(\leq) b_n} \sqrt{n} \tau e^{\sqrt{n} \tau (T_n - b_n)} dF_n = \]
\begin{equation*}
= \frac{1}{\sqrt{n\tau}} e^{nC + \sqrt{n\tau}s_{\alpha} + \frac{3}{6\tau^2}(1-Z^2)}(1+o(1))I_n
\end{equation*}
where

\begin{equation*}
I_n = \int_{T_n(\Omega, b_n)} \sqrt{n} \tau e^{\sqrt{n} \tau(T_n-b_n)} d\mathcal{P}.
\end{equation*}

It remains to estimate $I_n$.

By partial integration:

\begin{equation*}
I_n = \int_{0}^{\infty} P(T_n < b_n \text{ and } T_n(T_n-b_n) > \log Z) dz
\end{equation*}

where $T_n = \sqrt{n} \tau$. Hence:

\begin{equation*}
I_n = t_n \int_{0}^{\infty} [P_n(b_n) - P_n(b_n - \frac{\log Z}{t_n})] dz = A_n + B_n + C_n
\end{equation*}

where

\begin{equation*}
A_n = t_n \int_{0}^{\infty} [\Phi(b_n) - \Phi(b_n - \frac{\log Z}{t_n})] dz,
\end{equation*}

\begin{equation*}
B_n = t_n \int_{0}^{\infty} [R(b_n) - R(b_n - \frac{\log Z}{t_n})] dz
\end{equation*}

and

\begin{equation*}
C_n = o(1).
\end{equation*}

By partial integration again:

\begin{equation*}
A_n = t_n \int_{-\infty}^{b_n} e^{t_n(x-b_n)} \Phi(dx) = \Phi(b_n) t_n \left(1 - \frac{\Phi(T_n-b_n)}{t_n}\right) = \Phi'(s_a)(1+o(1)).
\end{equation*}

In order to estimate $B_n$ note that $\|R'\| < \infty$. Hence:
\[ B_n = \frac{x^3}{6t^2} \int_0^1 \frac{\log z}{t_n} R'(b_n - \frac{\log z}{t_n}) \, dz \]

so that

\[ |B_n| \leq \frac{\| R' \|}{6t^2} \int_0^1 (\log z) \, dz = \frac{\| R' \|}{6t^2} = o(1) . \]

Hence

\[ I_n = \delta'(s_\alpha) + o(1) \]

so that:

\[ 1 - \beta_n(\alpha) = \frac{1}{\sqrt{2\pi n}} \left( \frac{n(\sqrt{n} - t_s)}{2} + \frac{x^3}{6t^2} (1 - s^2 - \frac{3}{2}s^2) \right) (1 + o(1)) , \]

uniformly in \( \alpha \in [\alpha_0, \alpha_1] \).

Let us next drop the assumption that \( L_\mathbb{P}(Z) \) is not a lattice distribution. By Berry Esseèns inequality:

\[ F_n(b_n(\pm)) = \delta(b_n) + \frac{2\rho^3}{\sqrt{3/n}} \]

where \( \rho = \sqrt{E[Z^2 - \mu^2]} \) and \( |\theta| \), according to Zolotarev [ ], is \( \leq 0.82 \). It follows that \( b_n = s_\alpha + \frac{\theta c}{\sqrt{n}} \) for some constant \( c \) not depending on \( n \) where \( |\theta| \leq 1 \). The estimate of \( I_n \) may be written:

\[ I_n = A_n + \frac{2\rho^3}{\sqrt{3/n}} = \delta'(s_\alpha) + o(1) + \frac{2\rho^3}{\sqrt{3/n}} \]
It follows that there is a real constant $C$ so that:

$$1-\beta_n(\alpha) \leq \frac{C}{\sqrt{n}} e^{n C + \sqrt{n} \tau_\alpha}; \alpha \in [\alpha_0, \alpha_1]$$

References for section 5.


Asymptotic behaviour of deficiencies.
(References: Chernoff, LeCam, Torgersen)

Let \( X_1, X_2, \ldots \) and \( Y_1, Y_2, \ldots \) be two sequences of independent and identically distributed random variables. Suppose the distribution of \( X_j \) as well as of \( Y_j \) is determined up to the same unknown parameter \( \theta \). The purpose of this section is to compare, asymptotically, the statistical information stored in \( (X_1, \ldots, X_n) \) with the statistical information stored in \( (Y_1, \ldots, Y_n) \). We shall restrict ourselves to the case where there are only two possible values of \( \theta \).

An alternative formulation of this problem in terms of experiments is as follows. Let \( \mathcal{E} \) and \( \mathcal{F} \) be two dichotomies. What is the asymptotic behaviour of deficiencies of \( \mathcal{E}^n \) w.r.t. \( \mathcal{F}^n \) for large \( n \)?

It is not surprising that the idempotents play a particular role in this theory. There are, in the case of dichotomies, two types of idempotents. One consists of all totally uninformative dichotomies. These are the dichotomies \( (x, \mathcal{A}, P_1, P_2) \) where \( P_1 = P_2 \). We will denote this type by \( \mathcal{M}_a \). \( \mathcal{M}_a \) may also be interpreted as a particular choice within this type. The other type consists of all totally informative dichotomies. These are the dichotomies \( (x, \mathcal{A}, P_1, P_2) \) where \( P_1 \wedge P_2 = 0 \). This type will be denoted by \( \mathcal{M}_a \). \( \mathcal{M}_a \) may also be interpreted as a particular choice within this type.

All limits for dichotomies in this section are w.r.t. LeCam's pseudo distance \(^*\) \( \Delta \).

\(^*\) There is a difference between the pseudodistance \( \Delta \) in LeCam's paper [..] and the pseudodistance \( \Delta \) used here. \( \Delta(\mathcal{E}, \mathcal{F}) \) is in [..] defined as the sum of the deficiencies \( \delta(\mathcal{E}, \mathcal{F}) \) and \( \delta(\mathcal{F}, \mathcal{E}) \) while we let \( \Delta(\mathcal{E}, \mathcal{F}) \) denote the maximum of these deficiencies. These pseudodistances are clearly equivalent.
Clearly $\mathcal{C}^n \to \mathcal{M}a$ when $\mathcal{C} \not\sim \mathcal{M}i$. Hence $\Delta(\mathcal{C}^n, \mathcal{X}^n) \to 0$ when $\mathcal{C}, \mathcal{X} \not\sim \mathcal{M}i$. The immediate problem is then the rate of convergence. We know already that

$$\liminf_n \frac{n}{\sqrt{\Delta(\mathcal{C}^n, \mathcal{X}^n)}} > 0$$

when $\mathcal{C}, \mathcal{X} \not\sim \mathcal{M}i$. It is therefore natural to investigate the stability of $\frac{n}{\sqrt{\Delta(\mathcal{C}^n, \mathcal{X}^n)}}$ for large values of $n$.

Some notations we will use are:

Let $\mathcal{C} = (\chi, \mathcal{A}, P_1, P_2)$ denote any dichotomy. Then:

$\beta(\mathcal{C}) = \beta G(\alpha)$ is the power of the most powerful level $\alpha$ test for testing $P_1$ against $P_2$; $0 \leq \alpha \leq 1$. It is occasionally convenient to put $\beta(\mathcal{C}) = 1$ when $\alpha > 1$.

$B(\mathcal{C}) = B G(\lambda)$ is the minimum Baye's risk for the problem of testing $P_1$ against $P_2$ with the 0-1 loss function when the prior assigns probability $\lambda$ to $P_2$.

$\beta$ and $B$ are both characteristic of $\mathcal{C}$ in the sense that they, separately, describe $\mathcal{C}$ up to equivalence.

$\beta(\mathcal{C})$ is, and may be any, concave function on $[0, 1]$ such that $\beta(0^+) = \beta(0)$ and $\beta(1) = 1$.

$B(\mathcal{C})$ is, and may be any, concave function on $[0, 1]$ such that:

$$0 \leq B(\lambda) \leq (1-\lambda) \wedge \lambda$$

The dichotomy $\mathcal{C}$ is $(\epsilon_1, \epsilon_2)$ deficient w.r.t. if and only if:

$$\epsilon_1(1-\lambda)+\epsilon_2\lambda \geq 2[B(\lambda)-B(\lambda)]$$
In particular:

\[ \varepsilon(\mathcal{E}, \mathcal{F}) = 2\sup_{\lambda} |E_\mathcal{E}(\lambda) - E_\mathcal{F}(\lambda)| \]

and

\[ \Delta(\mathcal{E}, \mathcal{F}) = 2\sup_{\lambda} |E_\mathcal{E}(\lambda) - E_\mathcal{F}(\lambda)| \]

If we put \( \mathcal{F} = M_1 \) above then we get:

\[ \Delta(\mathcal{E}, M_1) = 1 - 2E_\mathcal{E}(\frac{1}{2}) \]

It is known, Torgersen [ , ] that \( \mathcal{E} \) is \( \varepsilon \)-deficient w.r.t. \( \mathcal{F} \) if and only if:

\[ \beta_\mathcal{E}(a + \frac{\varepsilon_1}{2}) + \frac{\varepsilon_2}{2} \geq \varepsilon_\mathcal{E}(a) ; a \geq 0 \]

This imply, in particular, that \( \Delta(\mathcal{E}, \mathcal{F}) \) is the Levy diagonal distance between \( \beta_\mathcal{E} \) and \( \beta_\mathcal{F} \). Hence:

\[ \Delta(\mathcal{E}, M_1) = \sup_{a} \beta(\varepsilon - a) \]

and

\[ \Delta(\mathcal{E}, M_\alpha) = 2\alpha_0 \text{ where } \alpha_0 \text{ is the unique solution of equation:} \]

\[ \alpha + \beta(\alpha) = 1 \]

The following inequality will be useful

Theorem 1

The distances \( \Delta(\mathcal{E}, M_\alpha) \) and \( \Delta(\mathcal{E}, M_1) \) satisfies the inequalities

\[ 1 - \Delta(\mathcal{E}, M_1) \leq \Delta(\mathcal{E}, M_\alpha) \leq \frac{1 - \Delta(\mathcal{E}, M_1)}{2 - \Delta(\mathcal{E}, M_1)} \]
Proof: The left inequality follows from the triangle inequality by:

\[ 1 = \Delta(M_i, Ma) \leq \Delta(M_i, \xi) + \Delta(\xi, Ma). \]

In proving the right inequality we will utilize the geometric interpretation of \( \Delta \). Let us write \( \delta_i = \Delta(\xi, M_i) \) and \( \delta_a = \Delta(\xi, Ma) \). Then \( \frac{\delta_1}{\sqrt{2}} \) is the maximal diagonal distance between the curve described by \( \beta \) and the segment \( \{(a, a) : 0 \leq a \leq 1\} \) while \( \frac{\delta_2}{\sqrt{2}} \) is the diagonal distance between this curve and the point \((0,1)\). Here is a picture of the situation:
The diagonal \([1,0),(0,1)\] of the unit square intersects the segments \([0,0),(0,1)\] and \([0,0),(1-\delta_i,1)\] in the same point \((x,y) = ((1-\delta_i)/(2-\delta_i),1/(2-\delta_i))\). Let \(\alpha \in [0,1]\) satisfy \(\beta(\alpha) = \alpha + \delta_i\). (There is at least one in \((0,1)\) with this property.) Suppose first that \(\alpha \leq x\). Then, since the points on the segment

\[[(0,\delta_i),(1,1)]\]

with abscissas \(\alpha\) and 1 both are below the graph of \(\beta\) and \(\beta\) is concave, \(\beta(x) \geq y\). The same argument applied to the segment \([0,0),(1-\delta_i,1)\]

yields \(\beta(x) \geq y\) when \(\alpha \geq x\). It follows that the point \((x,y)\) is below the graph of \(\beta\). Hence

\[
\frac{\delta_a}{\sqrt{2}} \leq \text{distance } ((0,1),(x,y)) = \sqrt{2} x
\]

i.e.

\[
\delta_a \leq 2 \cdot \frac{1-\delta_i}{2-\delta_i}
\]

**Corollary**

\[1-\Delta(\mathcal{E},\mathcal{M}i) \leq c(\mathcal{E})\]

and

\[\Delta(\mathcal{E},\mathcal{M}a) \leq 2 \frac{c(\mathcal{E})}{1+c(\mathcal{E})}\]

where

\[c(\mathcal{E}) = \inf \mathcal{E}(t)\]

---

**Proof:** \[1-\Delta(\mathcal{E},\mathcal{M}i) = 2\mathcal{E}(\frac{1}{2}) \leq 2 \inf_{0 < t < 1} (\frac{1}{2})^{1-t} (\frac{1}{2})^t \mathcal{E}(\frac{1}{2}) = c(\mathcal{E})\]

Hence:
\[ \Delta(\xi, \mathcal{M}a) \leq 2 \frac{1 - \Delta(\xi, \mathcal{M}i)}{2 - \Delta(\xi, \mathcal{M}i)} \leq 2 \frac{c(\xi)}{1 + c(\xi)}. \]

Put, for any experiment \( \xi \), \( c(\xi) = \inf \{ \log(t) : 0 < t < 1 \} \), Then our first convergence result is:

**Theorem 2**

For any dichotomy \( \mathcal{D} \):

\[
\lim_n \sqrt[n]{1 - \Delta(\mathcal{D}, \xi^n)} = c(\xi)
\]

and

\[
\lim_n \sqrt[n]{\Delta(\mathcal{D}, \xi^n)} = c(\xi)
\]

**Proof:** The first limit follows directly from Chernoff's results as they are formulated in section 4, and the simple relationship between \( B_\mathcal{D}(\xi) \) and \( \Delta(\xi, \mathcal{M}i) \). Thus

\[
n\sqrt[n]{1 - \Delta(\xi, \xi^n)} = n\sqrt[n]{2 \log (\frac{1}{n})} \to c(\xi)
\]

By the corollary:

\[
n\sqrt[n]{\Delta(\xi^n, \mathcal{M}a)} \leq n\sqrt[n]{2 c(\xi^n)} = n\sqrt[n]{\frac{c(\xi^n)}{1 + c(\xi^n)}} \to c(\xi)
\]

Hence

\[
\limsup_n \sqrt[n]{\Delta(\xi^n, \mathcal{M}a)} \leq c(\xi)
\]

By the theorem:

\[
\liminf_n \sqrt[n]{\Delta(\xi^n, \mathcal{M}a)} \geq \lim_n \sqrt[n]{1 - \Delta(\xi^n, \mathcal{M}i)} = c(\xi)
\]
It follows that:

\[
\limsup_n \sqrt{n} \Delta (\mathcal{G}^n, \mathcal{F}^n) \leq \limsup_n \sqrt{n} (\mathcal{G}^n, \mathcal{M}a + \Delta (\mathcal{M} a, \mathcal{F}^n)) = c(\mathcal{G}) \lor c(\mathcal{F}).
\]

On the other hand:

\[
\liminf_n \sqrt{n} \Delta (\mathcal{G}^n, \mathcal{F}^n) \geq \liminf_n \sqrt{n} \Delta (\mathcal{G}^n, \mathcal{M}a) - (\mathcal{F}^n, \mathcal{M}a)
\]

\[
= c(\mathcal{G}) \lor c(\mathcal{F}) \text{ provided } c(\mathcal{G}) \neq c(\mathcal{F}).
\]

Hence:

\[
\lim_n \Delta (\mathcal{G}^n, \mathcal{F}^n) = c(\mathcal{G}) \lor c(\mathcal{F}) \text{ provided } c(\mathcal{G}) \neq c(\mathcal{F}).
\]

It remains to consider the case \( c(\mathcal{G}) = c(\mathcal{F}) \).

**Example**

Let \( \mathcal{G} \) and \( \mathcal{F} \) be given by, respectively, the matrices:

\[
\begin{array}{c|c|c|c}
\mathcal{G} & 0 & 1 & 2 \\
\hline
0 & 1 & 1-\xi & \xi & 0 \\
1 & 2 & 0 & \xi & 1-\xi \\
\end{array}
\]

\[
\begin{array}{c|c|c|c}
\mathcal{G} & 0 & 1 & 2 \\
\hline
0 & 1 & 1-\eta & \eta & 0 \\
1 & 2 & \xi & 1-\xi \\
\end{array}
\]

where \( \eta > \xi \)

Then \( \mathcal{G}^n \) and \( \mathcal{F}^n \) are given by the matrices

\[
\begin{array}{c|c|c|c}
\mathcal{G}^n & 0 & 1 & 2 \\
\hline
0 & 1 & 1-\xi^n & \xi^n & 0 \\
1 & 2 & 0 & \xi^n & 1-\xi^n \\
\end{array}
\]

\[
\begin{array}{c|c|c|c}
\mathcal{F}^n & 0 & 1 & 2 \\
\hline
0 & 1 & 1-\eta^n & \eta^n & 0 \\
1 & 2 & 0 & \eta^n & 1-\eta^n \\
\end{array}
\]
Simple computations yield $c(\mathcal{E}) = c(\mathcal{F}) = \xi$ and

$$\Delta(\mathcal{E}^n, \mathcal{F}^n) = 2 \frac{\xi^n \eta^n}{\xi^n + \eta^n} - \xi^n ; \quad n = 1, 2, \ldots$$

Hence

$$\sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} - c(\mathcal{E})$$

The problem of convergence of $\sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)}$ is settled by

**Theorem 3**

Let $\mathcal{E}$ and $\mathcal{F}$ be any pair of non-equivalent dichotomies.

Then

$$\sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} - c(\mathcal{E}) \lor c(\mathcal{F}) \text{ as } n \to \infty.$$  

Remark: This shows the impossibility of approximating an experiment $\mathcal{E}$ by a hopefully more manageable experiment $\mathcal{F} \not\sim \mathcal{E}$ so that $\sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} = 0$.

By the remarks immediately before the example the convergence is already established provided $c(\mathcal{E}) \not= c(\mathcal{F})$.

We saw also that

$$\limsup_n \sqrt[n]{\Delta(\mathcal{E}^n, \mathcal{F}^n)} = c(\mathcal{E}) \lor c(\mathcal{F}) \text{ in any case.}$$

The proof in the case $c(\mathcal{E}) = c(\mathcal{F})$ is based on the following inequality:

Let $\mathcal{E}$, $\mathcal{F}$ and $\mathcal{G}$ be dominated experiments. Then

$$\Delta(\mathcal{E}, \mathcal{F}) \geq \Delta(\mathcal{E} \times \mathcal{G}, \mathcal{F} \times \mathcal{G})$$

It follows that the operation of multiplication with a fixed experiment $\mathcal{G}$ is a contraction. In particular:

$$\Delta(\mathcal{E}^n, \mathcal{F}^n) \geq \Delta(\mathcal{E}^n \times \mathcal{F}^n, \mathcal{E}^n \times \mathcal{F}^n).$$
where $\overline{\mathcal{E}} = \mathcal{E} \times \mathcal{F}$ and $\overline{\mathcal{H}} = \mathcal{H} \times \mathcal{F}$. If $\mathcal{F}$ contains little information then $\overline{\mathcal{E}}$ and $\overline{\mathcal{H}}$ are good approximations to, respectively, $\mathcal{E}$ and $\mathcal{F}$. Furthermore $\mathcal{F}$ may, as we shall see, be chosen so that $\overline{\mathcal{E}}$ and $\overline{\mathcal{H}}$ have desirable properties which $\mathcal{E}$ and $\mathcal{F}$ may not have.

We will also utilize the fact that the logarithm of a Hellinger transform is convex. Furthermore it is strictly convex provided it is not the Hellinger transform of any dichotomy of the form:

$$
\begin{array}{c|ccc}
\theta \backslash \mathcal{X} & 0 & 1 & 2 \\
\hline
1 & 1-\xi & \xi & 0 \\
2 & 0 & \eta & 1-\eta \\
\end{array}
$$

Note also that a Hellinger transform which is not the zero function is bounded away from 0 on $]0,1[$.

**Proof of the theorem:** We will use letters $L$, $M$ and $H$ with or without affixes to denote Hellinger transforms of dichotomies. We will also use self-explaining notations as $\Delta(L,M)$, $c(L)$ and $c(M)$. The logarithm of a number $x$ (function $f$) will be denoted by $\log(x(f))$.

By the remarks above and the symmetry of $\Delta$ the proof will be completed by proving:

**Statement:**

\[ \liminf_{n} \sqrt[n]{\Delta(L^n,M^n)} \geq c(L) \text{ when } L \neq M \]

and $L$ and $M$ are both bounded away from zero.
Proof of the statement: We will first show that we may, without loss of generality, assume that \( L'(0+) = M'(0+) = -L'(1-) = -M'(1-) = \infty \). Suppose we have proved the statement in this case and consider a general pair \((L, M)\) satisfying the conditions of the statement. Let \( H \) be a Hellinger transform of a dichotomy, so that \( H'(0+) = -H'(1-) = \infty \) [As an example put \( H(t) = \int_0^1 (1-x)^{1-t} x^t s(x) \, dx \) where \( s(x) = \tilde{s}(x) / \int_0^1 \tilde{s}(x) \, dx \) and
\[
\tilde{s}(x) = \begin{cases} 
  x^{-1} [\log x]^{-2} & \text{when } 0 < x < 1/4 \\
  0 & \text{when } 1/4 \leq x \leq 3/4 \\
  (1-x)^{-1} [\log(1-x)]^{-2} & \text{when } 3/4 \leq x < 1.
\end{cases}
\]

Let for each \( N = 1, 2, \ldots \), \( H_N = 1 - \frac{1}{N} H \) denote the Hellinger transform of the dichotomy \( (1-\frac{1}{N} \cup_i + \frac{1}{N} \cap) \). Put \( L_N = LH_N \) and \( M_N = MH_N \). Then \( L_N \) converges uniformly to \( L \) as \( N \to \infty \). Hence \( c(L_N) \to c(L) \). Furthermore, assuming \( L \neq 0 \) and \( M \neq 0 \),
\[
L_N'(0+) = M_N'(0+) = -L_N'(1-) = -M_N'(1-) = \infty .
\]

By assumption
\[
\liminf_{n \to \infty} \sqrt[n]{\Delta(L^N, M^N)} \geq \liminf_{n \to \infty} \sqrt[n]{\Delta(L_N^N, M_N^N)} \geq c(L_N) - c(L) \text{ as } N \to \infty
\]

It follows that we may, without loss of generality assume that
\[
L'(0+) = M'(0+) = -L'(1-) = -M'(1-) = \infty.
\]
Then \( L \) and \( M \) obtains minima in, respectively, \( t_L \in ]0,1[ \) and \( t_M \in ]0,1[ \). Choose \( a \geq 0 \) and put \( L_a(t) = e^{-at}L(t) \).
and \( M_a(t) = e^{-at}M(t) \). \( t \mapsto e^{-at} \) is the Hellinger transform of the dichotomy

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>( 0 )</th>
<th>( 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>( e^{-a} )</td>
<td>( 1 - e^{-a} )</td>
</tr>
</tbody>
</table>

It follows that \( I_a \) and \( M_a \) are different Hellinger transforms such that \( I_a'(0+) = M_a'(0+) = -I_a'(1-) = -M_a'(1-) = \infty \). Let \( t_{L,a} \) and \( t_{M,a} \) denote respectively, the points where \( I_a \) and \( M_a \) obtains minimum. Then

\[
\begin{align*}
  t_{L,a} &= [\tilde{L}']^{-1}(a) \quad \text{and} \quad t_{M,a} = [\tilde{M}']^{-1}(a).
\end{align*}
\]

Let us next argue that we may assume \( t_{L} \neq t_{M} \). Suppose first that \( L = M \). Then there is a real constant \( b \) so that \( \tilde{L} = \tilde{M} + b \) i.e. \( L = e^bM \). By assumption \( L \neq M \). Hence \( b \neq 0 \) so that \( c(L) = e^b c(M) \neq c(M) \) and we are through. If \( \tilde{L}' \neq \tilde{M}' \) then, since analytical functions are involved here, \( t_{L,a} \neq t_{M,a} \) when \( a \) is sufficiently small. Assuming we have a proof for the case \( t_{L} \neq t_{M} \) we get, provided \( a \) is sufficiently small:

\[
\liminf_{n \to \infty} \sqrt[n]{ \Delta(L^n, M^n) } \geq \liminf_{n \to \infty} \sqrt[n]{ \Delta(L^n_a, M^n_a) } \geq c(L_a) - c(L)
\]

as \( a \downarrow 0 \). It remains therefore to consider the case where \( t_{L} \neq t_{M} \).

By Lagrange's remainder formula:

\[
\tilde{L}(t) = \tilde{c}(L) + \frac{1}{2}(t-t_L)^2 \tilde{L}''(t) \quad \text{where} \quad t \in (t_L, t)
\]

Now

\[
\tilde{L}_a(t) = -at + \tilde{L}(t)
\]

so that

\[
\tilde{c}(I_a) = -at_L + \tilde{c}(L) + \frac{1}{2}(t_L, a - t_L)^2 \tilde{L}''(t_L, a).
\]
Furthermore \( (t_L, a-t_L)/a = ([\tilde{L}']^{-1}(a) - [\tilde{L}']^{-1}(0))/a \) converges, as \( a \downarrow 0 \), to the derivative of \([\tilde{L}']^{-1}\) in 0.

Hence
\[
\frac{\tilde{c}(L_a) - \tilde{c}(L)}{a} = -t_L, a + \frac{a}{2} \left( \frac{t_L, a-t_L}{2} \right) 2 \tilde{L}''(t_L, a) \rightarrow -t_L \text{ as } a \downarrow 0.
\]

Similarly
\[
\frac{\tilde{c}(M_a) - \tilde{c}(M)}{a} \rightarrow -t_M \text{ as } a \downarrow 0.
\]

It follows that \( \tilde{c}(L_a) = \log c(L_a) \neq \log c(M_a) = \tilde{c}(M_a) \) when \( a \) is sufficiently small. Hence, once more,
\[
\liminf_n \underline{\Delta}(L^n, M^n) \geq \liminf_n \underline{\Delta}(L^n_a, M^n_a) = c(L_a) - c(L) \text{ as } a \downarrow 0.
\]

References for section 6:

A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations.
AMS 23, 493-507.

Sufficiency and approximate sufficiency.
AMS 35, 1419-1455.

Comparison of experiments when the parameter space is finite.
References


L'information en statistique mathématique et dans la théorie des communications. Thèse, Faculté des Sciences de l'Université de Paris, June


[14] LeCam, L., 1964  
Sufficiency and approximate sufficiency. AMS 35, 1419-1455.

Théorie asymptotique de la decision statistique. Seminaire de Mathématique Superieures. Presses de L'Université de Montréal.

[16] Osipov, L.V. and Petrov, V.V., 1967  
On the estimation of the remainder term in the central limit theorem. Teoriya Veroyat. i. Primenen 12, vyp. 322-329.

[17] Petrov, V.V., 1955  


Comparison of experiments when the parameter space is finite. Zeit. f. Wahrscheinlichkeitstheorie u.v. Gebiete 16, 219-249.
