10.1 Binary Hypothesis Testing

Two possible distributions on a space $\mathcal{X}$

$$\begin{aligned} H_0 &: X \sim P \\ H_1 &: X \sim Q \end{aligned}$$

Where under hypothesis H_0 (the null hypothesis) X is distributed according to P, and under H_1 (the alternative hypothesis) X is distributed according to Q. A *test* between two distributions chooses either H_0 or H_1 based on an observation of X

- Deterministic test: $f : \mathcal{X} \to \{0, 1\}$
- Randomized test: $P_{Z|X} : \mathcal{X} \to \{0, 1\}$, so that $P_{Z|X}(0|x) \in [0, 1]$.

Let $Z = 0$ denote that the test chooses P, and $Z = 1$ when the test chooses Q.

Remark: This setting is called "testing simple hypothesis against simple hypothesis". Simple here refers to the fact that under each hypothesis there is only one distribution that could generate the data. Composite hypothesis is when $X \sim P$ and P is only known to belong to some class of distributions.

10.1.1 Performance Metrics

In order to determine the "effectiveness" of a test, we look at two metrics. Let $\pi_{i|j}$ denote the probability of the test choosing i when the correct hypothesis is j. With this

$$\begin{aligned} \alpha &= \pi_{0|0} = P[Z = 0] \quad \text{(Probability of success given } H_0 \text{ true)} \\ \beta &= \pi_{0|1} = Q[Z = 0] \quad \text{(Probability of error given } H_1 \text{ true)} \end{aligned}$$

Remark: $P[Z = 0]$ is a slight abuse of notation, more accurately $P[Z = 0] = \sum_{x \in \mathcal{X}} P(x) P_{Z|X}(0|x) = \mathbb{E}_{X \sim P_X}[1 - f(x)]$. Also, the choice of these two metrics to judge the test is not unique, we can use many other pairs from $\{\pi_{0|0}, \pi_{0|1}, \pi_{1|0}, \pi_{1|1}\}$.

So for any test $P_{Z|X}$ there is an associated (α, β). There are a few ways to determine the "best test"

- Bayesian: Assume prior distributions $\mathbb{P}[H_0] = \pi_0$ and $\mathbb{P}[H_1] = \pi_1$, minimize the expected error

$$P_b^* = \min_{\text{tests}} \pi_0 \pi_{1|0} + \pi_1 \pi_{0|1}$$

- Minimax: Assume there is a prior distribution but it is unknown, so choose the test that preforms the best for the worst case priors

$$P_m^* = \min_{\text{tests}} \max_{\pi_0} \pi_0 \pi_{1|0} + \pi_1 \pi_{0|1}$$

- Neyman-Pearson: Minimize error β subject to success probability at least α.

In this course, the Neyman-Pearson formulation will play a vital role.

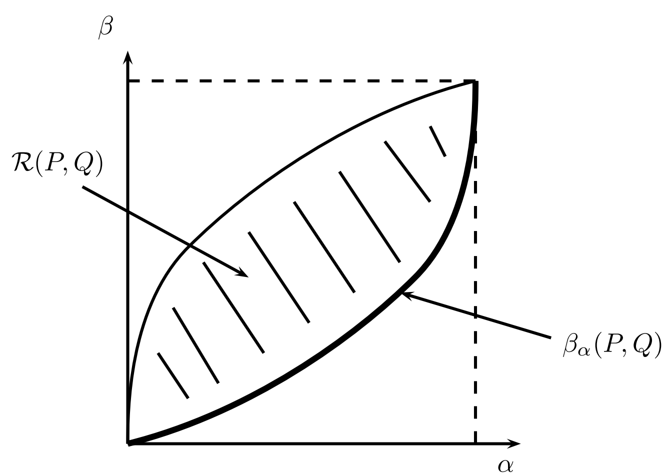
10.2 Neyman-Pearson formulation

Definition 10.1. Given that we require $P[Z=0] \geq \alpha$,

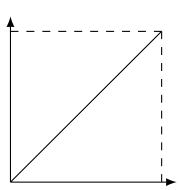
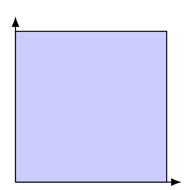
$$\beta_\alpha(P,Q) \triangleq \inf_{P[Z=0]\geq\alpha} Q[Z=0]$$

Definition 10.2. Given (P,Q), the region of achievable points for all randomized tests is

$$\mathcal{R}(P,Q) = \bigcup_{P_{Z|X}} \{(P[Z=0], Q[Z=0])\} \subset [0,1]^2 \tag{10.1}$$



Remark 10.1. This region encodes a lot of useful information about the relationship between P and Q. For example,[1]

$P = Q \Leftrightarrow \mathcal{R}(P,Q) =$  $P \perp Q \Leftrightarrow \mathcal{R}(P,Q) =$ 

Moreover, $\mathrm{TV}(P,Q)$ = maximal length of vertical line intersecting the lower half of $\mathcal{R}(P,Q)$ (HW).

Theorem 10.1 (Properties of $\mathcal{R}(P,Q)$)**.**

1. *$\mathcal{R}(P,Q)$ is a closed, convex subset of $[0,1]^2$.*

2. *$\mathcal{R}(P,Q)$ contains the diagonal.*

[1] Recall that P is mutually singular w.r.t. Q, denoted by $P \perp Q$, if $P[E]=0$ and $Q[E]=1$ for some E.

3. *Symmetry:* $(\alpha, \beta) \in \mathcal{R}(P, Q) \Leftrightarrow (1-\alpha, 1-\beta) \in \mathcal{R}(P, Q)$.

Proof. 1. For convexity, suppose $(\alpha_0, \beta_0), (\alpha_1, \beta_1) \in \mathcal{R}(P, Q)$, then each specifies a test $P_{Z_0|X}, P_{Z_1|X}$ respectively. Randomize between these two test to get the test $\lambda P_{Z_0|X} + \bar{\lambda} P_{Z_1|X}$ for $\lambda \in [0,1]$, which achieves the point $(\lambda\alpha_0 + \bar{\lambda}\alpha_1, \lambda\beta_0 + \bar{\lambda}\beta_1) \in \mathcal{R}(P, Q)$.

Closedness will follow from the explicit determination of all boundary points via Neyman-Pearson Lemma – see Remark 10.2. In more complicated situations (e.g. in testing against composite hypothesis) simple explicit solutions similar to Neyman-Pearson Lemma are not available but closedness of the region can frequently be argued still. The basic reason is that the collection of functions $\{g : \mathcal{X} \to [0,1]\}$ forms a weakly-compact set and hence its image under a linear functional $g \mapsto (\int g dP, \int g dQ)$ is closed.

2. Test by blindly flipping a coin, i.e., let $Z \sim \text{Bern}(1-\alpha) \perp\!\!\!\perp X$. This achieves the point (α, α).

3. If $(\alpha, \beta) \in \mathcal{R}(P, Q)$, then form the test that chooses P whenever $P_{Z|X}$ choses Q, and chooses Q whenever $P_{Z|X}$ choses P, which gives $(1-\alpha, 1-\beta) \in \mathcal{R}(P, Q)$.

$\square$

The region $\mathcal{R}(P, Q)$ consists of the operating points of all randomized tests, which include deterministic tests as special cases. The achievable region of deterministic tests are denoted by

$$\mathcal{R}_{\text{det}}(P, Q) = \bigcup_E \{(P(E), Q(E)\}. \tag{10.2}$$

One might wonder the relationship between these two regions. It turns out that $\mathcal{R}(P, Q)$ is given by the closed convex hull of $\mathcal{R}_{\text{det}}(P, Q)$.

We first recall a couple of notations:

- Closure: $\mathbf{cl}(E) \triangleq$ the smallest closed set containing E.
- Convex hull: $\mathbf{co}(E) \triangleq$ the smallest convex set containing $E = \{\sum_{i=1}^n \alpha_i x_i : \alpha_i \geq 0, \sum_{i=1}^n \alpha_i = 1, x_i \in E, n \in \mathbb{N}\}$. A useful example: if $(f(x), g(x)) \in E, \forall x$, then $(\mathbb{E}[f(X)], \mathbb{E}[g(X)]) \in \mathbf{cl}(\mathbf{co}(E))$.

Theorem 10.2 (Randomized test v.s. deterministic tests)**.**

$$\mathcal{R}(P, Q) = \mathbf{cl}(\mathbf{co}(\mathcal{R}_{\text{det}}(P, Q))).$$

Consequently, if P and Q are on a finite alphabet $\mathcal{X}$, then $\mathcal{R}(P, Q)$ is a polygon of at most $2^{|\mathcal{X}|}$ vertices.

Proof. "$\supset$": Comparing (10.1) and (10.2), by definition, $\mathcal{R}(P, Q) \supset \mathcal{R}_{\text{det}}(P, Q))$. By Theorem 10.1, $\mathcal{R}(P, Q)$ is closed convex, and we are done with the $\supset$ direction.

"$\subset$": Given any randomized test $P_{Z|X}$, put $g(x) = P_{Z=0|X=x}$. Then g is a measurable function. Moreover,

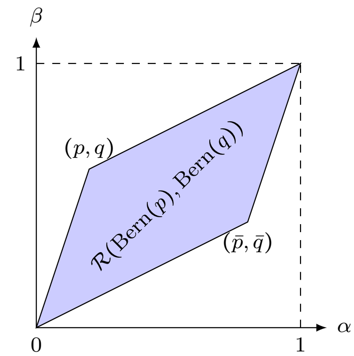
$$P[Z=0] = \sum_x g(x) P(x) = \mathbb{E}_P[g(X)] = \int_0^1 P[g(X) \geq t] \mathrm{d}t$$

$$Q[Z=0] = \sum_x g(x) Q(x) = \mathbb{E}_Q[g(X)] = \int_0^1 Q[g(X) \geq t] \mathrm{d}t$$

where we applied the formula $E[U] = \int \mathbb{P}[U \ge t]\,\mathrm{d}t$ for $U \ge 0$. Therefore the point $(P[Z=0], Q[Z=0]) \in \mathcal{R}$ is a mixture of points $(P[g(X) \ge t], Q[g(X) \ge t]) \in \mathcal{R}_{\text{det}}$, averaged according to t uniformly distributed on the unit interval. Hence $\mathcal{R} \subset \mathbf{cl}(\mathbf{co}(\mathcal{R}_{\text{det}}))$.

The last claim follows because there are at most $2^{|\mathcal{X}|}$ subsets in (10.2). $\square$

Example: Testing Bern(p) versus Bern(q), $p < \frac{1}{2} < q$. Using Theorem 10.2, note that there are $2^2 = 4$ events $E = \varnothing, \{0\}, \{1\}, \{0,1\}$. Then



10.3 Likelihood ratio tests

Definition 10.3. The log likelihood ratio (LLR) is $F = \log \frac{\mathrm{d}P}{\mathrm{d}Q} : \mathcal{X} \to \mathbb{R} \cup \{\pm\infty\}$. The likelihood ratio test (LRT) with threshold $\tau \in \mathbb{R}$ is $\mathbf{1}\{\log \frac{\mathrm{d}P}{\mathrm{d}Q} \le \tau\}$. Formally, we assume that $dP = p(x)d\mu$ and $dQ = q(x)d\mu$ (one can take $\mu = P + Q$, for example) and set

$$F(x) \triangleq \begin{cases} \log \frac{p(x)}{q(x)}, & p(x) > 0, q(x) > 0 \\ +\infty, & p(x) > 0, q(x) = 0 \\ -\infty, & p(x) = 0, q(x) > 0 \\ n/a, & p(x) = 0, q(x) = 0 \end{cases}$$

Notes:

- LRT is a deterministic test. The intuition is that upon observing x, if $\frac{Q(x)}{P(x)}$ exceeds a certain threshold, suggesting Q is more likely, one should reject the null hypothesis and declare Q.
- The rationale for defining extended values $\pm\infty$ of $F(x)$ are the following observations:

$$\begin{aligned} \forall x, \forall \tau \in \mathbb{R}: \qquad & (p(x) - \exp\{\tau\}q(x))1\{F(x) > \tau\} \ge 0 \\ & (p(x) - \exp\{\tau\}q(x))1\{F(x) \ge \tau\} \ge 0 \\ & (q(x) - \exp\{-\tau\}p(x))1\{F(x) < \tau\} \ge 0 \\ & (q(x) - \exp\{-\tau\}p(x))1\{F(x) \le \tau\} \ge 0 \end{aligned}$$

  This leads to the following useful consequence: For any $g \ge 0$ and any $\tau \in \mathbb{R}$ (note: $\tau = \pm\infty$ is excluded) we have

$$\mathbb{E}_P[g(X)1\{F \ge \tau\}] \ge \exp\{\tau\} \cdot \mathbb{E}_Q[g(X)1\{F \ge \tau\}] \tag{10.3}$$
$$\mathbb{E}_Q[g(X)1\{F \le \tau\}] \ge \exp\{-\tau\} \cdot \mathbb{E}_P[g(X)1\{F \le \tau\}] \tag{10.4}$$

Below, these and similar inequalities are only checked for the cases of F not taking extended values, but from this remark it should be clear how to treat the general case.

- Another useful observation:

$$Q[F = +\infty] = P[F = -\infty] = 0\,. \tag{10.5}$$

Theorem 10.3.

1. *F is a sufficient statistic for testing H_0 vs H_1.*
2. *For discrete alphabet $\mathcal{X}$ and when $Q \ll P$ we have*

$$Q[F = f] = \exp(-f)P[F = f] \qquad \forall f \in \mathbb{R} \cup \{+\infty\}$$

More generally, we have for any $g : \mathbb{R} \cup \{\pm\infty\} \to \mathbb{R}$

$$\mathbb{E}_Q[g(F)] = g(-\infty)Q[F = -\infty] + \mathbb{E}_P[\exp\{-F\}g(F)] \tag{10.6}$$
$$\mathbb{E}_P[g(F)] = g(+\infty)P[F = +\infty] + \mathbb{E}_Q[\exp\{F\}g(F)] \tag{10.7}$$

Proof. **(2)**

$$\begin{aligned} Q_F(f) &= \sum_{\mathcal{X}} Q(x)\mathbf{1}\{\log \frac{P(x)}{Q(x)} = f\} = \sum_{\mathcal{X}} Q(x)\mathbf{1}\{e^f Q(x) = P(x)\} \\ &= e^{-f} \sum_{\mathcal{X}} P(x)\mathbf{1}\{\log \frac{P(x)}{Q(x)} = f\} = e^{-f} P_F(f) \end{aligned}$$

To prove the general version (10.6), note that

$$\begin{aligned} \mathbb{E}_Q[g(F)] &= \int_{\{-\infty < F(x) < \infty\}} d\mu\, q(x) g(F(x)) + g(-\infty)Q[F = -\infty] && (10.8) \\ &= \int_{\{-\infty < F(x) < \infty\}} d\mu\, p(x) \exp\{-F(x)\} g(F(x)) + g(-\infty)Q[F = -\infty] && (10.9) \\ &= \mathbb{E}_P[\exp\{-F\}g(F)] + g(-\infty)Q[F = -\infty]\,, && (10.10) \end{aligned}$$

where we used (10.5) to justify restriction to finite values of F.

(1) To show F is a s.s, we need to show $P_{X|F} = Q_{X|F}$. For the discrete case we have:

$$\begin{aligned} P_{X|F}(x|f) &= \frac{P_X(x)P_{F|X}(f|x)}{P_F(f)} = \frac{P(x)\mathbf{1}\{\frac{P(x)}{Q(x)} = e^f\}}{P_F(f)} = \frac{e^f Q(x)\mathbf{1}\{\frac{P(x)}{Q(x)} = e^f\}}{P_F(f)} \\ &= \frac{Q_{XF}(xf)}{e^{-f}P_F(f)} \overset{(2)}{=} \frac{Q_{XF}}{Q_F} = Q_{X|F}(x|f). \end{aligned}$$
□

The general argument is done similarly to the proof of (10.6).

From Theorem 10.2 we know that to obtain the achievable region $\mathcal{R}(P,Q)$, one can iterate over all subsets and compute the region $\mathcal{R}_{\det}(P,Q)$ first, then take its closed convex hull. But this is a formidable task if the alphabet is huge or infinite. But we know that the LLR $\log \frac{\mathrm{d}P}{\mathrm{d}Q}$ is a sufficient statistic. Next we give bounds to the region $\mathcal{R}(P,Q)$ in terms of the statistics of $\log \frac{\mathrm{d}P}{\mathrm{d}Q}$. As usual, there are two types of statements:

- Converse (outer bounds): any point in $\mathcal{R}(P,Q)$ must satisfy ...
- Achievability (inner bounds): the following point belong to $\mathcal{R}(P,Q)$...

10.4 Converse bounds on $\mathcal{R}(P, Q)$

Theorem 10.4 (Weak Converse). *$\forall(\alpha, \beta) \in \mathcal{R}(P, Q)$,*

$$d(\alpha\|\beta) \le D(P\|Q)$$
$$d(\beta\|\alpha) \le D(Q\|P)$$

where $d(\cdot\|\cdot)$ is the binary divergence.

Proof. Use data processing with $P_{Z|X}$. □

Lemma 10.1 (Deterministic tests). *$\forall E,\ \forall \gamma > 0:\ P[E] - \gamma Q[E] \le P\Big[\log \frac{dP}{dQ} > \log \gamma\Big]$*

Proof. (Discrete version)

$$\begin{aligned} P[E] - \gamma Q[E] &= \sum_{x \in E} p(x) - \gamma q(x) \le \sum_{x \in E} (p(x) - \gamma q(x)) \mathbf{1}_{\{p(x) > \gamma q(x)\}} \\ &= P\Big[\log \frac{dP}{dQ} > \log \gamma, X \in E\Big] - \gamma Q\Big[\log \frac{dP}{dQ} > \log \gamma, X \in E\Big] \le P\Big[\log \frac{dP}{dQ} > \log \gamma\Big]. \end{aligned}$$

(General version) WLOG, suppose $P, Q \ll \mu$ for some measure μ (since we can always take $\mu = P + Q$). Then $dP = p(x) d\mu$, $dQ = q(x) d\mu$. Then

$$\begin{aligned} P[E] - \gamma Q[E] &= \int_E d\mu (p(x) - \gamma q(x)) \le \int_E d\mu (p(x) - \gamma q(x)) \mathbf{1}_{\{p(x) > \gamma q(x)\}} \\ &= P\Big[\log \frac{dP}{dQ} > \log \gamma, X \in E\Big] - Q\Big[\log \frac{dP}{dQ} > \log \gamma, X \in E\Big] \le P\Big[\log \frac{dP}{dQ} > \log \gamma\Big]. \end{aligned}$$

where the second line follows from $\frac{p}{q} = \frac{\frac{dP}{d\mu}}{\frac{dQ}{d\mu}} = \frac{dP}{dQ}$.

[So we see that the only difference between the discrete and the general case is that the counting measure is replaced by some other measure μ.] □

Note: In this case, we do not need $P \ll Q$, since $\pm\infty$ is a reasonable and meaningful value for the log likelihood ratio.

Lemma 10.2 (Randomized tests). *$P[Z = 0] - \gamma Q[Z = 0] \le P\Big[\log \frac{dP}{dQ} > \log \gamma\Big]$.*

Proof. Almost identical to the proof of the previous Lemma 10.1:

$$\begin{aligned} P[Z = 0] - \gamma Q[Z = 0] &= \sum_x P_{Z|X}(0|x)(p(x) - \gamma q(x)) \le \sum_x P_{Z|X}(0|x)(p(x) - \gamma q(x)) \mathbf{1}_{\{p(x) > \gamma q(x)\}} \\ &= P\Big[\log \frac{dP}{dQ} > \log \gamma, Z = 0\Big] - Q\Big[\log \frac{dP}{dQ} > \log \gamma, Z = 0\Big] \\ &\le P\Big[\log \frac{dP}{dQ} > \log \gamma\Big]. \end{aligned}$$
□

Theorem 10.5 (Strong Converse). *$\forall(\alpha, \beta) \in \mathcal{R}(P, Q), \forall \gamma > 0$,*

$$\alpha - \gamma \beta \le P\Big[\log \frac{dP}{dQ} > \log \gamma\Big] \tag{10.11}$$

$$\beta - \frac{1}{\gamma} \alpha \le Q\Big[\log \frac{dP}{dQ} < \log \gamma\Big] \tag{10.12}$$

Proof. Apply Lemma 10.2 to (P, Q, γ) and $(Q, P, 1/\gamma)$. □

Note: Theorem 10.5 provides an outer bound for the region $\mathcal{R}(P,Q)$ in terms of half-spaces. To see this, suppose one fixes $\gamma > 0$ and looks at the line $\alpha - \gamma\beta = c$ and slowing increases c from zero, there is going to be a maximal c, say c^*, at which point the line touches the lower boundary of the region. Then (10.11) says that c^* cannot exceed $P[\log \frac{dP}{dQ} > \log\gamma]$. Hence $\mathcal{R}$ must lie to the left of the line. Similarly, (10.12) provides bounds for the upper boundary. Altogether Theorem 10.5 states that $\mathcal{R}(P,Q)$ is contained in the intersection of a collection of half-spaces indexed by γ.
Note: To apply the strong converse Theorem 10.5, we need to know the CDF of the LLR, whereas to apply the weak converse Theorem 10.4 we need only to know the expectation of the LLR, i.e., divergence.

10.5 Achievability bounds on $\mathcal{R}(P,Q)$

Since we know that the set $\mathcal{R}(P,Q)$ is convex, it is natural to try to find all of its supporting lines (hyperplanes), as it is well known that closed convex set equals the intersection of the halfspaces correposponding to all supporting hyperplanes. So thus, we are naturally lead to solving the problem

$$\max\{\alpha - t\beta : (\alpha,\beta) \in \mathcal{R}(P,Q)\}.$$

This can be done rather simply:

$$\alpha^* - t\beta^* = \max_{(\alpha,\beta)\in\mathcal{R}} (\alpha - t\beta) = \max_{P_{Z|X}} \sum_{x\in\mathcal{X}} (P(x) - tQ(x)) P_{Z|X}(0|x) = \sum_{x\in\mathcal{X}} |P(x) - tQ(x)|^+$$

where the last equality follows from the fact that we are free to choose $P_{Z|X}(0|x)$, and the best choice is obvious:

$$P_{Z|X}(0|x) = 1\left\{\log\frac{P(x)}{Q(x)} \ge \log t\right\}.$$

Thus, we have shown that all supporting hyperplanes are parameterized by LLR-tests. This completely recovers the region $\mathcal{R}(P,Q)$ except for the points corresponding to the faces (linear pieces) of the region. To be precise, we state the following result.

Theorem 10.6 (Neyman-Pearson Lemma). *"LRT is optimal": For any α, β_α is attained by the following test:*

$$P_{Z|X}(0|x) = \begin{cases} 1 & \log\frac{dP}{dQ} > \tau \\ \lambda & \log\frac{dP}{dQ} = \tau \\ 0 & \log\frac{dP}{dQ} < \tau \end{cases} \tag{10.13}$$

where $\tau \in \mathbb{R}$ and $\lambda \in [0,1]$ are the unique solutions to $\alpha = P[\log\frac{dP}{dQ} > \tau] + \lambda P[\log\frac{dP}{dQ} = \tau]$.

Proof of Theorem 10.6. Let $t = \exp(\tau)$. Given any test $P_{Z|X}$, let $g(x) = P_{Z|X}(0|x) \in [0,1]$. We want to show that

$$\alpha = P[Z=0] = \mathbb{E}_P[g(X)] = P\Big[\frac{dP}{dQ} > t\Big] + \lambda P\Big[\frac{dP}{dQ} = t\Big] \tag{10.14}$$

$$\Rightarrow \beta = Q[Z=0] = \mathbb{E}_Q[g(X)] \overset{\text{goal}}{\ge} Q\Big[\frac{dP}{dQ} > t\Big] + \lambda Q\Big[\frac{dP}{dQ} = t\Big] \tag{10.15}$$

Using the simple fact that $\mathbb{E}_Q[f(X)\mathbf{1}_{\{\frac{dP}{dQ}\le t\}}] \ge t^{-1}\mathbb{E}_P[f(X)\mathbf{1}_{\{\frac{dP}{dQ}\le t\}}]$ for any $f \ge 0$ twice, we have

$$\begin{aligned}
\beta &= \mathbb{E}_Q[g(X)\mathbf{1}_{\{\frac{dP}{dQ}\le t\}}] + \mathbb{E}_Q[g(X)\mathbf{1}_{\{\frac{dP}{dQ}> t\}}] \\
&\ge \frac{1}{t}\underbrace{\mathbb{E}_P[g(X)\mathbf{1}_{\{\frac{dP}{dQ}\le t\}}]} + \mathbb{E}_Q[g(X)\mathbf{1}_{\{\frac{dP}{dQ}> t\}}] \\
&\overset{(10.14)}{=} \frac{1}{t}\underbrace{\Big(\mathbb{E}_P[(1-g(X))\mathbf{1}_{\{\frac{dP}{dQ}> t\}}] + \lambda P\Big[\frac{dP}{dQ} = t\Big]\Big)} + \mathbb{E}_Q[g(X)\mathbf{1}_{\{\frac{dP}{dQ}> t\}}] \\
&\ge \mathbb{E}_Q[(1-g(X))\mathbf{1}_{\{\frac{dP}{dQ}> t\}}] + \lambda Q\Big[\frac{dP}{dQ} = t\Big] + \mathbb{E}_Q[g(X)\mathbf{1}_{\{\frac{dP}{dQ}> t\}}] \\
&= Q\Big[\frac{dP}{dQ} > t\Big] + \lambda Q\Big[\frac{dP}{dQ} = t\Big]. \qquad \square
\end{aligned}$$

Remark 10.2. As a consequence of the Neyman-Pearson lemma, all the points on the boundary of the region $\mathcal{R}(P,Q)$ are attainable. Therefore

$$\mathcal{R}(P,Q) = \{(\alpha,\beta) : \beta_\alpha \le \beta \le 1 - \beta_{1-\alpha}\}.$$

Since $\alpha \mapsto \beta_\alpha$ is convex on $[0,1]$, hence continuous, the region $\mathcal{R}(P,Q)$ is a closed convex set. Consequently, the infimum in the definition of β_α is in fact a minimum.

Furthermore, the lower half of the region $\mathcal{R}(P,Q)$ is the convex hull of the union of the following two sets:

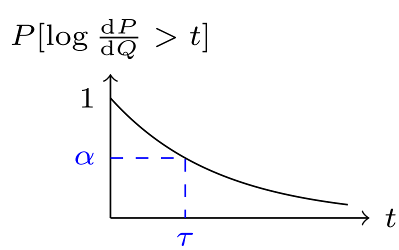
$$\begin{cases} \alpha = P\big[\log\frac{dP}{dQ} > \tau\big] \\ \beta = Q\big[\log\frac{dP}{dQ} > \tau\big] \end{cases} \qquad \tau \in \mathbb{R}\cup\{\pm\infty\}.$$

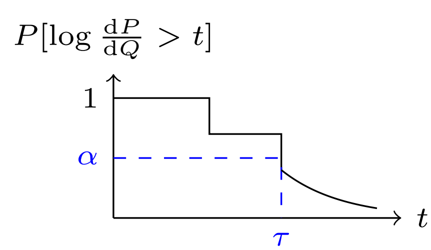
and

$$\begin{cases} \alpha = P\big[\log\frac{dP}{dQ} \ge \tau\big] \\ \beta = Q\big[\log\frac{dP}{dQ} \ge \tau\big] \end{cases} \qquad \tau \in \mathbb{R}\cup\{\pm\infty\}.$$

Therefore it does not lose optimality to restrict our attention on tests of the form $\mathbf{1}\{\log\frac{dP}{dQ} \ge \tau\}$ or $\mathbf{1}\{\log\frac{dP}{dQ} > \tau\}$.

Remark 10.3. The test (10.13) is related to LRT[2] as follows:





1. Left figure: If $\alpha = P[\log\frac{dP}{dQ} > \tau]$ for some τ, then $\lambda = 0$, and (10.13) becomes the LRT $Z = \mathbf{1}_{\{\log\frac{dP}{dQ}\le\tau\}}$.
2. Right figure: If $\alpha \ne P[\log\frac{dP}{dQ} > \tau]$ for any τ, then we have $\lambda \in (0,1)$, and (10.13) is equivalent to randomize over tests: $Z = \mathbf{1}_{\{\log\frac{dP}{dQ}\le\tau\}}$ with probability $\bar\lambda$ or $\mathbf{1}_{\{\log\frac{dP}{dQ}<\tau\}}$ with probability λ.

[2] Note that it so happens that in Definition 10.3 the LRT is defined with an $\le$ instead of $<$.

Corollary 10.1. *$\forall \tau \in \mathbb{R}$, there exists $(\alpha, \beta) \in \mathcal{R}(P, Q)$ s.t.*

$$\alpha = P\Big[\log \frac{\mathrm{d}P}{\mathrm{d}Q} > \tau\Big]$$
$$\beta \le \exp(-\tau) P\Big[\log \frac{\mathrm{d}P}{\mathrm{d}Q} > \tau\Big] \le \exp(-\tau)$$

Proof.

$$\begin{aligned} Q\Big[\log \frac{\mathrm{d}P}{\mathrm{d}Q} > \tau\Big] &= \sum Q(x) \mathbf{1}\Big\{\frac{P(x)}{Q(x)} > e^{\tau}\Big\} \\ &\le \sum P(x) e^{-\tau} \mathbf{1}\Big\{\frac{P(x)}{Q(x)} > e^{\tau}\Big\} = e^{-\tau} P\Big[\log \frac{\mathrm{d}P}{\mathrm{d}Q} > \tau\Big]. \end{aligned}$$ □

10.6 Asymptotics

Now we have many samples from the underlying distribution

$$H_0 : X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P$$
$$H_1 : X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} Q$$

We're interested in the asymptotics of the error probabilities $\pi_{0|1}$ and $\pi_{1|0}$. There are two main types of tests, both which the convergence rate to zero error is exponential.

1. Stein Regime: What is the best exponential rate of convergence for $\pi_{0|1}$ when $\pi_{1|0}$ has to be $\le \epsilon$?

$$\begin{cases} \pi_{1|0} \le \epsilon \\ \pi_{0|1} \to 0 \end{cases}$$

2. Chernoff Regime: What is the trade off between exponents of the convergence rates of $\pi_{1|0}$ and $\pi_{0|1}$ when we want both errors to go to 0?

$$\begin{cases} \pi_{1|0} \to 0 \\ \pi_{0|1} \to 0 \end{cases}$$