Chapter 3: Concentration of Measure:

(Shannon Entropy)

$$H(x) := -\sum_{x\in X} p(x)\log p(x)$$

(Functional)

$$\mathbb{H}_\phi(X) := \mathbb{E}[\phi(X)] - \phi(\mathbb{E}[X])$$

3.1

a) If $\phi(x) = H(x)$, then the outcome:

$$\mathbb{H}_\phi(x) := \mathbb{E}[\phi(x)] - \phi(\mathbb{E}[X])$$

$$= \mathbb{E}[H(x)] - H(\mathbb{E}[X])$$

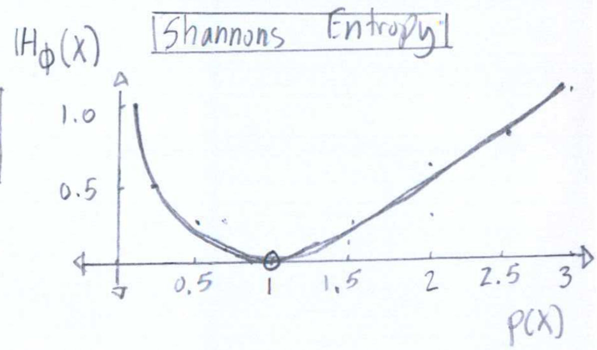
$$= -\mathbb{E}\Big[\sum_{x\in X} p(x)\log p(x)\Big] - \sum_{x\in X} p(\mathbb{E}(x))\log p(\mathbb{E}(x)) \qquad \boxed{p(x) = \frac{1}{X}}$$

$$= \frac{1}{|X|}\sum_{x\in X}\left(\log|X| - H(x)\right)$$

b) Uniform Distribution: $p(X) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & x<a \text{ or } x>a \end{cases}$

Shannon's Entropy maximized:

$$\frac{d}{dp}\mathbb{H}_\phi(x) = \frac{d}{dp}\left[p(x)\circ\log p(x) + \frac{1}{p(x)}\log\left(\frac{1}{p(x)}\right)\right]$$



$$p(x)^* = 1$$

$$= \frac{1}{b-a}$$

= Uniform distribution when $b=1$ and $a=0$.

c) If $Y = p(x)/q(x)$, Shannon's entropy becomes a divergence between two distributions.

$$\mathbb{H}_\phi(Y) = \mathbb{E}[\phi(Y)] - \phi(\mathbb{E}[X])$$

$$= -\mathbb{E}\Big[\sum_{x\in X} Y\log Y\Big] + \sum_{x\in X} Y(\mathbb{E}(x))\log Y(\mathbb{E}[X])$$

$$= 0 + \sum_{X\in X} Y(X)\log Y(X)$$

$$= +\sum_{x\in X}\frac{p(X)}{Q(X)}\log\frac{p(X)}{Q(X)}$$

3.2.

$$D(\mathbb{Q}\|\mathbb{P}) = \sum_{x\in X} p(x,y) \log \frac{P(x,y)}{Q(x,y)}$$

$$= \sum_{x\in X} p(x) \log \frac{p(x,y)}{Q(x,y)}$$

$$+ \sum_{x\in X} p(y|x) \log \frac{p(x,y)}{Q(x,y)}$$

$$= D(\mathbb{Q}_1, \mathbb{P}_1) + \sum_{x\in X} D(\mathbb{Q}_{x-1} \| \mathbb{P}_{x-1})$$

Joint Probability:
$P(x,y) = P(x) \circ P(y)$
or
$P(x,y) = P(x) \circ p(y|x)$

3.3. $\mathbb{H}(e^{\lambda X}) = \mathbb{E}[\lambda X e^{\lambda X}] - \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda X}]$

$$= \mathbb{E}[\lambda X e^{\lambda X}] - \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda X}] + \mathbb{E}[e^{\lambda X}] - \mathbb{E}[e^{\lambda X}]$$

If $\log \mathbb{E}[e^{\lambda X}] = \lambda t$, then an expansion appears:

$$\mathbb{H}(e^{\lambda X}) = \mathbb{E}[\lambda X e^{\lambda X}] - \lambda t \mathbb{E}[e^{\lambda X}] + e^{\lambda t} - \mathbb{E}[e^{\lambda X}]$$

$$= \mathbb{E}[(e^{\lambda(X-t)} - 1 + \lambda(X-t)) e^{\lambda X}]$$

$$= \mathbb{E}[\psi(\lambda(X-t)) e^{\lambda X}] \quad \text{where } \psi(u) = e^{-u} - 1 + u$$

$$= \inf_{t\in\mathbb{R}} \mathbb{E}[\psi(\lambda(X-t)) e^{\lambda X}]$$

3.4. a) $\mathbb{H}(e^{\lambda(X+c)}) = \mathbb{E}[\lambda(X+c) e^{\lambda(X+c)}] - \mathbb{E}[e^{\lambda(X+c)}] \log \mathbb{E}[e^{\lambda(X+c)}]$

$$= e^{\lambda c}\left[\mathbb{E}[\lambda X e^{\lambda X}] + \mathbb{E}[\lambda c e^{\lambda X}] - \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda X}] + \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda c}]\right]$$

$$= e^{\lambda c}\left[\mathbb{E}[\lambda X e^{\lambda X}] - \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda X}]\right]$$

$$= e^{\lambda c} \circ \mathbb{H}(e^{\lambda X})$$

b) When $\mathbb{H}(e^{\lambda X}) \le \frac{1}{2} \sigma^2 \lambda^2 \phi_X(\lambda)$

$$e^{\lambda c} \mathbb{H}(e^{\lambda X}) \le \frac{e^{\lambda c}}{2} \circ \sigma^2 \circ \lambda^2 \circ \phi_X(\lambda)$$

$$\mathbb{H}(e^{\lambda(X+c)}) \le \frac{e^{\lambda c}}{2} \circ \sigma^2 \circ \lambda^2 \circ \phi_X(\lambda)$$

3.5. $\phi(u) = u\log u - u$

$$\mathbb{H}_\phi(u) = \mathbb{E}[\phi(u)] - \phi(\mathbb{E}[u])$$

$$\mathbb{H}_\phi(e^{\lambda X}) = \mathbb{E}[e^{\lambda X} \circ \log e^{\lambda X} - e^{\lambda X}] - \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda X}] + \mathbb{E}[e^{\lambda X}]$$

$$= \mathbb{E}[\lambda X \cdot e^{\lambda X}] - \mathbb{E}[e^{\lambda X}] \log \mathbb{E}[e^{\lambda X}]$$

$$= \mathbb{H}(e^{\lambda X})$$

(Bernstein Entropy Bound)

$$\mathbb{H}(e^{\lambda X}) \le \lambda^2[b\phi_X'(\lambda) + \phi_X(\lambda)(\sigma^2 - b\mathbb{E}[X])] \quad \text{for all } \lambda \in [0, 1/b)$$

$$\mathbb{E}[e^{\lambda(X - \mathbb{E}[X])}] \le e^{\frac{\sigma^2\lambda^2}{(1-b\lambda)}} \quad \text{for all } \lambda \in [0, 1/b)$$

3.6.

a) $\mathbb{H}(e^{\lambda X}) \le \lambda^2 e^{\lambda \mathbb{E}[X]}[b\phi_{\bar{X}}'(\lambda) + \phi_X(\lambda)\sigma^2]$

$$\le \lambda^2[b[\phi_X'(\lambda) - \phi_X(\lambda)\mathbb{E}[X]] + \phi_X(\lambda)\sigma^2]$$

$$\mathbb{H}(e^{\lambda \bar{X}}) \le \lambda^2 e^{-\lambda\mathbb{E}[X]}[b\phi_X'(\lambda) + \phi_X(\lambda)[\sigma^2 - b\mathbb{E}[X]]]$$

$$\le \lambda^2[b[\phi_{\bar{X}}'(\lambda) + \phi_{\bar{X}}(\lambda)\mathbb{E}[X]] + \phi_{\bar{X}}(\lambda)[\sigma^2 - b\mathbb{E}[X]]]$$

b) When $\bar{X} = X/b$ and $\bar{\sigma}^2 = \sigma^2/b^2$, the equation moves:

$$\mathbb{H}(e^{\lambda X/b}) \le \frac{\lambda^2}{b^2}[b\mathbb{E}[b\tfrac{X}{b}e^{\lambda X/b}] + \sigma^2\mathbb{E}[e^{\lambda X/b}]]$$

$$\mathbb{H}(e^{\lambda b\bar{X}}) \le \lambda^2 b^2[b\mathbb{E}[\tfrac{X}{b}e^{\lambda b X/b}] + \frac{\sigma^2}{b^2}\mathbb{E}[e^{\lambda b X/b}]]$$

3.7.

$$\mathbb{H}(e^{\lambda X}) = \lambda^2\sigma^2\phi_X(\lambda) - \frac{1}{2}\lambda^2\sigma^2\phi_X(\lambda)$$

$$\le \frac{\lambda^2\sigma^2}{2}\phi_X(\lambda) \quad \text{where } \sigma = (b-a)/2$$

3.8. Exponential Distribution: $p_\theta(y) = h(y)e^{\langle\theta, T(y)\rangle - \Phi(\theta)}$

where $T: y \to \mathbb{R}^d$

$h(y) = \text{constant}$

Log Normalization: $\Phi(\theta) = \log\int_y e^{\langle\theta, T(y)\rangle} \circ h(y) \circ \mu \circ dy$

Lipschitz Parameter: $\|\nabla\Phi(\theta) - \nabla\Phi(\theta')\|_2 \le L\|\theta - \theta'\|_2$

a) $X = \langle v, T(Y) \rangle$

$$\mathbb{H}(e^{\lambda X}) = \mathbb{H}(e^{\lambda \langle v, T(y) \rangle})$$
$$= \lambda \phi'_{\langle v, T(y) \rangle}(\lambda) - \phi_{\langle v, T(y) \rangle}(\lambda) \log(\phi_{\langle v, T(y) \rangle}(\lambda))$$
$$= \lambda \phi_{\langle v, T(y) \rangle}(\lambda) \log'(\phi_{\langle t, T(v) \rangle}(\lambda)) - \phi_{\langle v, T(y) \rangle}(\lambda) \log \phi_{\langle v, T(y) \rangle}(\lambda)$$
$$= [\lambda \log'(\phi_{\langle v, T(y) \rangle}(\lambda)) - \log(\phi_{\langle v, T(y) \rangle}(\lambda)] \phi_{\langle v, T(y) \rangle}(\lambda)$$
$$= \left[\int_0^\lambda \lambda \log'(\phi_{\langle v, T(y) \rangle}(\lambda) - \log(\phi_{\langle v, T(y) \rangle}(\lambda)) \circ t\right] dt \circ \phi_{\langle v, T(y) \rangle}(\lambda)$$
$$= \lambda^2 \left[\int_0^\lambda \frac{\log'(\phi_{\langle v, T(y) \rangle}(\lambda)}{\lambda} - \frac{\log(\phi_{\langle v, T(y) \rangle}(\lambda)) \circ t}{\lambda^2}\right] dt \circ \phi_{\langle v, T(y) \rangle}(\lambda)$$
$$\leq L \lambda^2 \phi_{\langle v, T(y) \rangle}(\lambda) \quad \text{where } L = \frac{\| \nabla \Phi(\theta) - \nabla \Phi(\theta') \|}{\| \theta - \theta' \|}$$
$$\leq \frac{\lambda^2 \sigma^2}{2} \phi_{\langle v, T(y) \rangle}(\lambda) \quad \text{if } \frac{\sigma^2}{2} = \sqrt{2L}$$

b) i) Gaussian Distribution: $Y \sim N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\mathbb{H}(e^{\lambda X}) = \lambda \phi'_Y(x) - \phi_Y(x) \log \phi_Y(x)$$
$$= -\frac{\lambda^2 (x-\mu)^2}{\sigma^2} \phi_Y(x) + \phi_Y(x) \frac{\lambda^2 (x-\mu)^2}{2\sigma^2}$$
$$= \frac{\lambda^2 (x-\mu)^2}{2\sigma^2} \phi_Y(x)$$
$$\leq \lambda^2 L \phi_Y(x) \quad \text{where } L = \frac{(x-\mu)^2}{2\sigma^2}$$

ii) Bernoulli Distribution: $Y = p^x (1-p)^{1-x}$

$$\mathbb{H}(e^{\lambda Y}) = \lambda \phi'_Y(x) - \phi_Y(x) \log \phi_Y(x)$$
$$= \lambda^2 \left[\int_0^\lambda \frac{\log' \phi_Y(x)}{\lambda} - \frac{\log \phi_Y(x) \circ x}{\lambda^2}\right] dx \; \phi_Y(x)$$
$$\leq \lambda^2 \circ L \circ \phi_Y(x)$$
$$\text{where } L = \frac{\| \nabla \Phi(x) - \nabla \Phi(x') \|}{\| x - x' \|}$$

3.9. $\mathbb{H}(e^{\lambda f(x)}) = \lambda \phi'(x) - \phi'(x) \log \phi(x)$

$$\mathbb{H}(e^{\lambda f(x)}) - \mathbb{E}[g(x)e^{\lambda f(x) - g(x)}] = \lambda \phi'(x) - \phi(x) \log \phi(x) - \mathbb{E}[g(x) e^{\lambda f(x) - g(x)}]$$

$$= \lambda f(x) \phi(x) - \phi(x) \log \phi(x) - \mathbb{E}[g(x) e^{\lambda f(x) - g(x)}]$$

$$= \phi(x)[\lambda f(x) - g(x)] e^{-g(x)} - \phi(x) \log \phi(x)$$

If $\lambda f(x) - g(x) = \log \phi(x)$, then an equality appears:

$$\mathbb{H}(e^{\lambda f(x)}) - \mathbb{E}[g(x)e^{\lambda f(x) - g(x)}] = [1 - e^{-g(x)}] \phi(x) \log \phi(x)$$

When $g(x) \leq 1$ or $\lambda \leq \dfrac{\log g(x)}{f(x)}$, the supremum holds:

$$\mathbb{H}(e^{\lambda f(x)}) = \sup_g [\mathbb{E}[g(x) e^{\lambda f(x)}] \mid \mathbb{E}[e^{g(x)}] \leq 1]$$

Note: As two datasets separate from one dataset, an additional term scales above or below the raw data

(Brunn-Minkowski Inequality)

$$[\mu(A+B)]^{1/n} \geq [\mu(A)]^{1/n} + [\mu(B)]^{1/2}$$

where A, B are nonempty sets
$A + B := \{a + b \in \mathbb{R}^n \mid a \in A, b \in B\}$
μ = Lebesgue Measure
$n \geq 1$

(Classical Isoperimetric Inequality)

$$\text{per}(s) \geq n \,\text{vol}(s)^{n-1/n} \circ \text{vol}(B_1)^{1/n}$$

where $\text{per}(s)$ = perimeter of set $S \subset \mathbb{R}^n$
$\text{vol}(B_1)$ = volume $B_1 \subset \mathbb{R}^n$

3.10.

a) $[\text{vol}(\lambda C + (1-\lambda)D)]^{1/n} \geq [\text{Vol}(\lambda C)]^{1/n} [\text{vol}((1-\lambda)D]^{1/n}$

$$= \lambda^n (1-\lambda)^n [\text{vol}(C)]^{1/n} [\text{vol}(D)]^{1/n}$$

$$\geq \lambda^n [\text{vol}(C)]^{1/n} + (1-\lambda)[\text{vol}(D)]^{1/n}$$

b) $\lambda^n \text{vol}(C)^{1/n} + (1-\lambda) \text{vol}(D)^{1/n} \geq [\text{Vol}(C)]^{\lambda/n} \circ [\text{vol}(D)]^{(1-\lambda)/n}$

c) $[\text{vol}(\lambda C + (1-\lambda)D)]^{1/n} = \text{vol}\left(\dfrac{\lambda A + (1-\lambda)B}{\lambda \,\text{vol}(A)^{1/n} + (1-\lambda)\text{vol}(B)^{1/n}}\right)^{1/n}$

$$= \text{vol}\left(\frac{\lambda A}{\lambda \,\text{vol}(A)^{1/n} + (1-\lambda)\text{vol}(B)^{1/n}} + \frac{(1-\lambda) B}{\lambda \,\text{vol}(A)^{1/n} + (1-\lambda)\,\text{vol}(B)^{1/n}} \right)^{1/n}$$

$$\geq \lambda^n [\text{vol}(C)]^{1/n} + (1-\lambda)[\text{vol}(D)]^{1/n}$$

3.11.

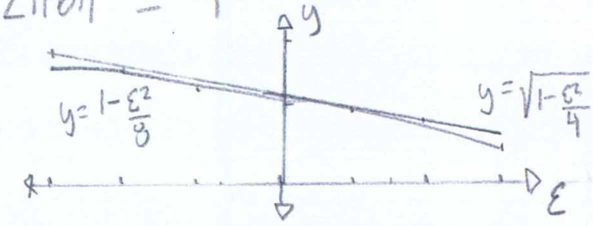
a) $B_2^n = \{ x \in \mathbb{R}^n \mid \|x\|_2 \leq 1 \}$

If $\|a-b\| = \varepsilon$, then

$$\|a+b\|^2 + \|a-b\|^2 = 2\|a\|^2 + 2\|b\|^2 \leq 4$$

$$\|a+b\|^2 + \varepsilon^2 \leq 4$$

$$\frac{1}{2}\|a+b\| \leq \sqrt{1 - \frac{\varepsilon^2}{4}} \sim 1 - \frac{\varepsilon^2}{8}$$



The maximum error on a Euclidean ball when data concentrates on the surface.

b) $\frac{\|a+b\|^n}{2^n} = \left[\text{vol}(A) \circ \text{vol}(A^{\varepsilon,c}) \right]^{n/2} = \mathbb{P}[A](1 - \mathbb{P}[A])^{n/2} \leq (1 - \frac{1}{8}\varepsilon^2)$

$$\mathbb{P}[A](1-\mathbb{P}[A]) \leq (1 - \frac{1}{8}\varepsilon^2)^{2n}$$

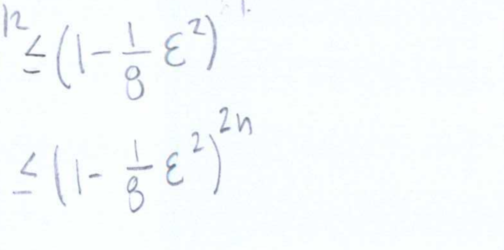
c) $\alpha_{\mathbb{P}(X,\mathbb{P})} = 1 - \mathbb{P}(A^\varepsilon) = \mathbb{P}((A^\varepsilon)^c) \leq 2(1 - \frac{1}{8}\varepsilon^2)^{2n} \leq 2 \cdot e^{-\frac{1}{4}n\varepsilon^2}$

This model functions for n-dimension spheroids with $n \geq 2$.

3.12 $Q \in S_+^{d \times d}$; Rademacher Chaos Variable : $X = \sum_{j=1}^{d} Q_{ij} \varepsilon_i \varepsilon_j$

a) $\mathbb{P}[X \geq (\sqrt{\text{Trace}(Q)} + t)^2] = \mathbb{P}[\sum Q_{ij} \varepsilon_i \varepsilon_j \geq (\sqrt{\text{Trace}(Q)} + t)^2]$

Page 44 shows an example about $(\sqrt{\text{Trace}(Q)} + t)^2 \leq 2\sqrt{\text{Trace}(Q)}^2 + 2t^2$.



A substitution list:

- $X = \sum Q^{1/2} \circ Q^{1/2} \varepsilon_i \cdot \varepsilon_j$
  $= \|M\|^2$
- $\sqrt{\text{Trace}(Q)} = 2\|M\|^2$
- $t^2 = 8\delta\|M\|$

$$\mathbb{P}[X \geq (\sqrt{\mathrm{Trace}(Q)}+t)^2] \leq \mathbb{P}[X \geq 2\sqrt{\mathrm{Trace}(Q)} + 2t^2]$$
$$\leq \mathbb{P}[\|M\|^2 \geq 4\|M\|^2 + 16\delta\|M\|]$$

A Hanson-Wright Inequality from Chapter 2:

$$\mathbb{P}[\|M\|^2 \geq 4\|M\|^2 + 16\delta\|M\|] \leq 2\exp\left[-\min\left(\frac{\delta^2}{4\|M\|^2+16\delta\|M\|}, \frac{\delta}{2\|M\|}\right)\right]$$

After backsubstitution,

$$\mathbb{P}[X \geq (\sqrt{\mathrm{Trace}\, Q}+t)^2] \leq 2\exp\left(\frac{-t^2}{16\delta\|Q\|}\right)$$

b. The other term from Hanson-Wright's Inequality where $\frac{\|M\|^2 - 4\|M\|}{16} = Y$:

$$\mathbb{P}[Y \geq \delta] \leq 2\exp\left(-\frac{\delta^2}{4\|M\|^2+16\delta\|M\|^2}\right)$$

(Wasserstein Distances)

$$W_p(Q, \mathbb{P}) = \sup_{\|f\|_{Lip} \leq 1}\left[\int f\,dQ - \int f\,d\mathbb{P}\right]$$

3.13. $$W_p(\mathbb{P}, Q) = \inf_{\mathbb{M}} \mathbb{M}[X \neq Y] = \sup_{\|f\|_{Lip} \leq 1}\left[\int f\,dQ - \int f\,d\mathbb{P}\right]$$
$$= \sup_{\|f\|_{Lip} \leq 1}\left[\int f \circ q\, d\mu - \int f \circ p\, d\mu\right]$$
$$= \sup_{\|f\|_{Lip} \leq 1}\left[\int f(q-p)\, d\mu\right]$$

3.14: Proposition 3.20:

$$W_p(Q, \mathbb{P}) \leq \sqrt{2\left(\sum_{k=1}^{n} \gamma_k\right) D(Q\,|\,\mathbb{P})}$$ where γ_k is a scalar for the univariate distribution

In this case, $n=2$ and $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$

a) $$W_p(Q, \mathbb{P}) = \sup_{\|f\|_{Lip} \leq 1}\left[\int f\,dQ - \int f\,d\mathbb{P}\right]$$
$$= \sup_{\|f\|_{Lip} \leq 1}\left[\int f(x_1, x_2)\, dQ - \int f(x_1, x_2)\, d\mathbb{P}\right]$$
$$\leq \sup_{\|f\|_{Lip} \leq 1}\left[\int f(x_1, x_2)\, dQ_{21}\, dQ_1 - \int f(x_1, x_2)\, d\mathbb{P}_2\, d\mathbb{P}_1\right]$$
$$\leq \sup_{\|f\|_{Lip} \leq 1}\left[\int\left[\int f(x_1, x_2)(dQ_2 - d\mathbb{P}_1)\right] dQ_1 + \int\left[\int f(x_1, x_2)\, d\mathbb{P}_2\right](dQ_1 - d\mathbb{P}_1)\right]$$

b) $$W_\rho(\mathbb{Q},\mathbb{P}) \leq \sup_{\|f\|\leq 1}\left\{\int\left[\int f(x_1,x_2)(d\mathbb{Q}_2-d\mathbb{P}_1)\right]d\mathbb{Q}_1+\int\left[\int f(x_1,x_2)\,d\mathbb{P}_2\right](d\mathbb{Q}_1-d\mathbb{P}_2)\right]$$

$$\leq \sqrt{2\gamma_2 D(\mathbb{Q}_2\|\mathbb{P}_2)} + \sqrt{2\gamma_1 D(\mathbb{Q}_1\|\mathbb{P}_1)}$$

$$\leq \left[\int\sqrt{2\gamma_2 D(\mathbb{Q}_2\|\mathbb{P}_2)}\,d\mathbb{Q}_1\right]+\sqrt{2\gamma_1 D(\mathbb{Q}_1\|\mathbb{P}_1)}$$

c) $$\leq \int\left[\sqrt{2\gamma_2 D(\mathbb{Q}_2\|\mathbb{P}_2)}+\sqrt{2\gamma_1 D(\mathbb{Q}_1\|\mathbb{P}_1)}\right]d\mathbb{Q}_1$$

$$\leq \int\sqrt{2(\gamma_1+\gamma_2)[D(\mathbb{Q}_0,\mathbb{P}_2)+D(\mathbb{Q}_1\|\mathbb{P}_1)}\,d\mathbb{Q}_1$$

$$\leq \sqrt{2(\gamma_1+\gamma_2)[\mathbb{E}[D(\mathbb{Q}_0,\mathbb{P}_2)]+D(\mathbb{Q}_1,\mathbb{P}_1)}$$

$$\leq \sqrt{2(\gamma_1+\gamma_2)\mathbb{D}(\mathbb{Q}\|\mathbb{P})}$$

$$\leq \sqrt{2\sum_{i=1}^{2}\gamma_i\,\mathbb{D}(\mathbb{Q}\|\mathbb{P})}$$

3.15.

a) $Y_k(X) := (X_1,\ldots, X_{k-1}, 0, X_{k+1}, X_n)$; $Z(X) = \max_{j=1,\ldots,m}\sum_{i=1}^{n} X_i^j$

$$Z(X)-Z(Y_k(X)) = Z(Y_{k,j\neq k}(X)) \geq 0$$

b) $$\mathbb{H}(e^{\lambda Z(X)}) \leq \mathbb{E}\left[\sum_{k=1}^{n}\mathbb{H}(e^{\lambda Z(X)}|X^{\backslash k})\right]$$

$$\leq \mathbb{E}\left[\sum_{k=1}^{n}\mathbb{E}[\psi(\lambda(Z(X)-Z(Y_k(X))))e^{\lambda Z(X)}|X^{\backslash k}]\right]$$

where Exercise 3.3 : $\mathbb{H}(e^{\lambda X}) = \inf_{t\in\mathbb{R}}\mathbb{E}[\psi(\lambda(X-t))e^{\lambda X}]$

c) From part a,

$$\lambda[Z(X)-Z(Y_k(X))] \geq 0$$

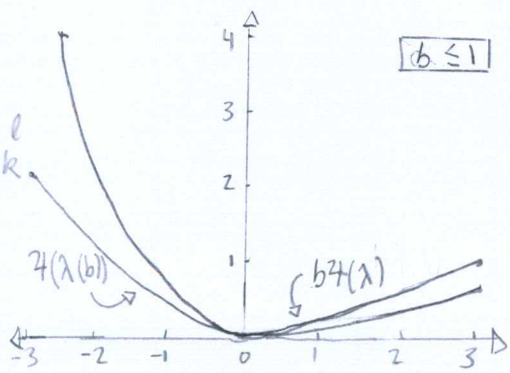
Also, $\sum_{l=1}^{M}[X\in\mathbb{A}]X_k^l$ is the entire X-set with X_k

$$\lambda\sum_{i=1}^{M}[X\in\mathbb{A}]X_k^l \geq \lambda[Z(X)-Z(Y_k(X))] \geq 0$$

d) When $\lambda(Z(X)-Z(Y_k(x)))\leq 1$, then

$$\psi(\lambda(Z(x)-Z(Y_k(X))) \leq \psi(\lambda)\sum_{l=1}^{M}[X\in\mathbb{A}]X_k^l$$

because $\psi(\lambda(b)) \leq b\,\psi(\lambda)$, but only for cases that b is less than or equal to one.



e) $$\mathbb{H}(e^{\lambda Z}) \le \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{E}\left[\psi(\lambda(Z(X) - Z(Y_k(X))))\, e^{\lambda Z(X)} \mid X^{\setminus k}\right]\right]$$

$$\le \mathbb{E}\left[\sum_{k=1}^{n} \mathbb{E}\left[\psi(\lambda)\left[\sum_{i=1}^{M} [X \in \mathbb{A}]\, X_k^{\ell}\right] e^{\lambda Z(X)}\right]\right]$$

$$\le \psi(\lambda) \sum_{k=1}^{n} \mathbb{E}\left[\sum_{\ell=1}^{M} [X \in \mathbb{A}]\, X_k^{\ell}\, e^{\lambda Z(X)}\right] = \psi(\lambda)\, \mathbb{E}\left[Z(X) e^{\lambda Z(X)}\right]$$

f) $$\mathbb{H}(e^{\lambda Z}) = \lambda \phi'(\lambda) - \phi(\lambda) \log \phi(\lambda) \le \psi(\lambda) \circ \phi'(\lambda)$$

$$\lambda (\log' \phi(\lambda)) - \log \phi(\lambda) \le \psi(\lambda) \log' \phi(\lambda)$$

$$\log' \phi(\lambda) \le \frac{1}{\lambda - \psi(\lambda)} \log \phi(\lambda)$$

$$\le \frac{e^{\lambda}}{e^{\lambda} - 1} \log \phi(\lambda) \quad \text{when} \quad \psi(\lambda) = e^{\lambda} - 1 + \lambda$$

(Bernstein Tail Bound)

$$\mathbb{P}[Z \ge \mathbb{E}[Z] + \delta] \le e^{\left(\frac{-n\delta^2}{c_1 \gamma^2 + c_2 b \delta}\right)}$$

3.16.

a) $$\mathbb{P}\left[Z \ge \mathbb{E}[Z] + \gamma \sqrt{\frac{c_1 t}{n}} + \frac{c_2 b t}{n}\right] = \mathbb{P}\left[Z \ge \mathbb{E}[Z] + \delta\right]$$

$$\delta = \gamma \sqrt{\frac{c_1 t}{n}} + \frac{c_2 b t}{n} \quad \text{where} \quad t = \frac{-n\delta^2}{c_1 \gamma^2 + c_2 b \delta}$$

$$\mathbb{P}\left[Z \ge \mathbb{E}[Z] + \gamma \sqrt{\frac{c_1 t}{n}} + \frac{c_2 b t}{n}\right] \le e^{-t}$$

b) $$\gamma^2 \le \sigma^2 + c_3 b \mathbb{E}[Z]$$

$$\mathbb{P}\left[Z \ge \mathbb{E}[Z] + \gamma \sqrt{\frac{c_1 t}{n}} + \frac{c_2 b t}{n}\right]$$

$$\le \mathbb{P}\left[Z \ge \mathbb{E}[Z] + \left(\sigma + \sqrt{c_3 b \mathbb{E}[Z] t}\right) \sqrt{\frac{c_1 t}{n}} + \frac{c_2 b t}{n}\right]$$

$$\le \mathbb{P}\left[Z \ge \mathbb{E}[Z] + \left(\sigma + \sqrt{2 c_3 b \mathbb{E}[Z] t}\right) \sqrt{\frac{c_1 t}{n}} + \frac{c_2 b t}{n}\right]$$

$$\le \mathbb{P}\left[Z \ge \mathbb{E}[Z] + \left(\sigma + \epsilon \mathbb{E}[Z] + \sqrt{\frac{c_1 c_3 b t}{2 \epsilon n}}\right) \sqrt{\frac{c_1 t}{n}} + \frac{c_2 b t}{n}\right]$$

$$\le \mathbb{P}\left[Z \ge (1+\epsilon) \mathbb{E}[Z] + \sigma \sqrt{\frac{c_1 t}{n}} + \left(c_2 + \frac{c_1 c_3}{2\epsilon}\right) \frac{b t}{n}\right]$$

$$\le e^{-t}$$