1. **Total variation distance.** Consider a measurable space $(\mathcal{X},\mathcal{F})$ and the set $\mathcal{M}(\mathcal{X})$ of probability distributions on $\mathcal{X}$. The total variation distance for $P,Q \in \mathcal{M}(\mathcal{X})$, is defined as
$$\mathrm{TV}(P,Q) \triangleq \frac{1}{2}\mathbb{E}_Q\left|\frac{dP}{dQ} - 1\right|.$$
   (a) Show that $\mathrm{TV}(P,Q) = 1 - \int_{\mathcal{X}} d(P \wedge Q)$.
   (b) Show that $\mathrm{TV} : \mathcal{M}(\mathcal{X}) \times \mathcal{M}(\mathcal{X}) \to \mathbb{R}_+$ is a metric on the space of probability distributions $\mathcal{M}(\mathcal{X})$.

2. **f-divergence.** Consider a measurable space $(\mathcal{X},\mathcal{F})$, the set $\mathcal{M}(\mathcal{X})$ of probability distributions on $\mathcal{X}$, and a map $f : (0,\infty) \to \mathbb{R}_+$. The f-divergence for any $P,Q \in \mathcal{M}(\mathcal{X})$ such that $P \ll Q$ is defined as
$$D_f(P\|Q) \triangleq \mathbb{E}_Q f\Big(\frac{dP}{dQ}\Big).$$
Consider $h : (\infty,0) \to \mathbb{R}_+$ such that $h(x) \triangleq f(x) + a(x-1)$ for some $a \in \mathbb{R}$ and all $x \in \mathcal{X}$. Show that $D_h = D_f$, i.e. functions differing in a linear term lead to the same f-divergence.

3. **χ^2 divergence.** Let $P,Q \in \mathcal{M}(\mathcal{X})$ Consider χ^2 divergence defined as $\chi^2(P\|Q) \triangleq \int_{\mathcal{X}} dQ\Big(\frac{dP}{dQ} - 1\Big)^2$. Show that $D(P\|Q) \leqslant \ln(1 + \chi^2(P\|Q))$.

4. **Hellinger distance.** Consider the Hellinger distance $H(P,Q) \triangleq \sqrt{H^2(P,Q)}$
   (a) Show that Hellinger distance $H : \mathcal{M}(\mathcal{X}) \times \mathcal{M}(\mathcal{X}) \to \mathbb{R}_+$ defines a metric on the space of probability distributions $\mathcal{M}(\mathcal{X})$.
   (b) Show that $P \mapsto H(P,Q)$ is not a convex map.

5. **Le Cam distance.** Consider Le Cam distance defined as $\mathrm{LC}(P,Q) \triangleq \frac{1}{2}\int_{\mathcal{X}} \frac{(dQ-dP)^2}{dQ+dP}$ for all $P,Q \in \mathcal{M}(\mathcal{X})$. Show that $\sqrt{\mathrm{LC}} : \mathcal{M}(\mathcal{X}) \times \mathcal{M}(\mathcal{X}) \to \mathbb{R}_+$ defines a metric on the space of probability distributions $\mathcal{M}(\mathcal{X})$.

6. **Jensen Shannon divergence.** Consider Jensen Shannon divergence defined as $\mathrm{JS}(P,Q) \triangleq D(P\|\frac{1}{2}(P+Q)) + D(Q\|\frac{1}{2}(P+Q))$ for all $P,Q \in \mathcal{M}(\mathcal{X})$. Show that $\sqrt{\mathrm{JS}} : \mathcal{M}(\mathcal{X}) \times \mathcal{M}(\mathcal{X}) \to \mathbb{R}_+$ defines a metric on the space of probability distributions $\mathcal{M}(\mathcal{X})$.

7. **KL Divergence for GLM**. Consider the Gaussian location model, where the parametrized distribution for observations is given by $P^\theta \triangleq \mathcal{N}(\theta,\Sigma)$.
   (a) Show that $J_F(\theta) = \Sigma^{-1}$.
   (b) Consider unconstrained parameter space $\Theta \subseteq \mathbb{R}^d$, and Gaussians $P^{\theta_i} \triangleq \mathcal{N}(\theta_i,\Sigma_i)$ for $i \in \{0,1\}$. Assuming $\det\Sigma_0 \neq 0$, show that
$$D(P^{\theta_0}\|P^{\theta_1}) = \frac{1}{2}(\theta_0-\theta_1)^T\Sigma_1^{-1}(\theta_0-\theta_1) + \frac{1}{2}\Big(\ln\det\Sigma_1 - \ln\det\Sigma_0 + \mathrm{tr}(\Sigma_1^{-1}\Sigma_0 - I_d)\Big).$$

8. **Local behavior of divergence of mixtures.** Let $\lambda \in [0,1], i \in \{0,1\}$ and $P_i,Q_i \in \mathcal{M}(\mathcal{X})$, to define mixture distribution $P_i^\lambda \triangleq \bar{\lambda}P_i + \lambda Q_i$. Show that under suitable technical conditions, the following equations hold
$$\frac{d}{d\lambda}\Big|_{\lambda=0} D(P_0^\lambda\|P_1) = \mathbb{E}_{Q_0}\ln\frac{dP_0}{dP_1} - D(P_0\|P_1),$$
$$\frac{d}{d\lambda}\Big|_{\lambda=0} D(P_1^\lambda\|P_0^\lambda) = \mathbb{E}_{Q_1}\ln\frac{dP_1}{dP_0} - D(P_1\|P_0) + \mathbb{E}_{P_1}\Big[1 - \frac{dQ_0}{dP_0}\Big].$$

9. **Fisher information for mixture of distributions.** Let $P,Q \in \mathcal{M}(\mathcal{X})$ and $\Theta \triangleq [0,1]$ and define $P^\theta \triangleq \theta P + \bar{\theta}Q \in \mathcal{M}(\mathcal{X})$. Show that $\lim_{\theta\downarrow 0} J_F(\theta) = \chi^2(P\|Q)$.