Sufficiency and Related Notions

35. Let $\tilde{m}$ be the Lebesgue measure restricted on $(1, \infty)$ and $\nu$ be the point mass measure at 0, that is, $\nu(\{0\}) = 1$.

(a) Use the above notation, we have

$$
P_X^\lambda(A) = P(X' > 1, X' \in A) + I_A(0)P(X' \leq 1)$$

$$= \int_A \lambda \exp(-\lambda x)d\tilde{m} + I_A(0)(1 - e^{-\lambda})$$

$$= \int_A \lambda \exp(-\lambda x)d\tilde{m} + \int_A (1 - e^{-\lambda})d\nu$$

$$= \int_A \lambda \exp(-\lambda x)I\{x \neq 0\}d(\tilde{m} + \nu) + \int_A (1 - e^{-\lambda})I\{x = 0\}d(\tilde{m} + \nu)$$

Thus $\frac{dP_X^\lambda}{d\mu}(x) = \lambda \exp(-\lambda x)I\{x \neq 0\} + (1 - e^{-\lambda})I\{x = 0\}$, a.s.$(\mu)$.

(b) Let $\delta_i = I\{X_i \neq 0\}$, one can see that $\delta_i = I\{X_i > 1\}$, a.e.$(\mu)$. The likelihood function is $\lambda^{\sum_i \delta_i} \exp\{-\lambda \sum_i \delta_i X_i\}(1 - e^{-\lambda})^{n - \sum_i \delta_i}$, which implies that $T(X) = (\sum_i \delta_i, \sum_i \delta_i X_i)$ is a sufficient statistic for this problem.

(c) Consider the ratio of $\prod_{i=1}^n \frac{dP_X^\lambda}{d\mu}(X_i)$ and $\prod_{i=1}^n \frac{dP_X^\lambda}{d\mu}(Y_i)$, which is

$$(1 - e^{-\lambda})^{\sum_i \delta_i^Y - \sum_i \delta_i^X} \lambda^{\sum_i \delta_i^Y} \exp\{\lambda(\sum_i \delta_i^Y Y_i - \delta_i^X X_i)\},$$

where $\delta_i^X, \delta_i^Y$ are the corresponding $\delta_i$'s to $X$ and $Y$ respectively. Clearly, this ratio independent of $\lambda$ implies that $T(X) = T(Y)$. Thus $T(X)$ is also minimal sufficient by Theorem 55.

36. Note that $I\{W = X'\} = I\{X' \leq Z\}$.

(a) In order to get the R-N derivative, consider the sets $A \times B$ where $A, B \in B$. Then $P(X \in A \times B) = P(X \in (\{0\} \cap A) \times B) + P(X \in (\{1\} \cap A) \times B) = I_A(0)P(X \in \{0\} \times B) + I_A(1)P(X \in \{1\} \times B) = I_A(0)P(X' > Z, Z \in B) + I_A(1)P(X' \leq Z, X' \in B) = I_A(0) \int_B \exp\{-(1 + \lambda)z\}dm(z) + I_A(1) \int_B \lambda \exp\{-(1 + \lambda)x'\}dm(x') = \int_{A \times B} \exp\{-(1 + \lambda)t_2\}(I_{\{t_1 = 0\}}(t_1) + \lambda I_{\{t_1 = 1\}}(t_1))dm(t_1, t_2)$. Thus

$$\frac{dP_X^\lambda}{d\mu}(t_1, t_2) = \exp\{-(1 + \lambda)t_2\}I\{t_1 = 0\} + \lambda I\{t_1 = 1\}I\{t_2 > 0\} = \lambda^{I\{t_1 = 1\}}\exp\{-(1 + \lambda)t_2\}I\{t_2 > 0\}$, a.e.$(\mu)$.

(b) Let $\delta_i = I\{w_i = x_i'\}$, then join density is

$$\lambda^{\sum_i \delta_i} \exp\{-\lambda \sum_i w_i\} I\{\min_i w_i > 0\} = \exp\{\log(\lambda) \sum_i \delta_i - \lambda \sum_i w_i\} I\{\min_i w_i > 0\}$$

A minimal sufficient statistic here would be $(\sum_i I\{W_i = X_i'\}, \sum_i W_i)$. Claim 72 says that this is truely a minimal sufficient statistic. As one can see that, this is an exponential family and $\Gamma_\lambda = \{ (\log \lambda, \lambda) | \lambda > 0 \}$ contains an open rectangle in $\mathbb{R}^2$ (We assume the parametric space $\Theta = \mathbb{R}^+$.)

1
37. The likelihood function is
\[ \frac{dP_\theta}{d\mu}(X) = \Lambda(\theta, X)f_\theta(X), \]
by factorization theorem, \( \Lambda(\theta, X) \) is sufficient. Moreover, \( \frac{dP_\theta}{d\mu}(x) = \frac{dP_\theta(y)f_\theta(x)}{f_\theta(y)}, \forall \theta \Rightarrow f_\theta(x) = f_\theta(y) \Rightarrow \Lambda(\theta, x) = \Lambda(\theta, y) \), thus \( \Lambda(\theta, X) \) is minimal sufficient by Theorem 55.

38. Denote \( \delta_i = I\{X_i > d\}, i = 1, \ldots, n \).
(a) Because \( T(X) \) is sufficient for \( X_1, \ldots, X_n \) that are iid \( P_\theta \), by factorization theorem (Theorem 50), there exists \( B \)-measurable function \( h \) and \( F \)-measurable function \( g_\theta \) such that \( \forall \theta \prod_i f_\theta(X_i) = g_\theta(T(X))h(X) \). Joint likelihood function of the truncated model is
\[ \prod_i \delta_i g_\theta(T(X))h(X)/(1 - F(d))^n = I\{\min_i X_i > d\}g_\theta(T(X))h(X)/(1 - F(d))^n, \]
thus \( (T(X), \min_i X_i) \) is sufficient.
(b) No. As a matter of fact, Theorem 55 can be a if and only if statement. One can see that as long as \( \min(\min_i X_i, \min_i Y_i) > d \) and \( T(X) = T(Y) \), the ratio of \( f_{\theta, d}(X) \) with respect to \( f_{\theta, d}(Y) \) would be independent of \( (\theta, d) \) due to the minimal sufficiency of \( T(X) \) for \( \theta \). But this does not say that \( \min_i X_i = \min_i Y_i \).

39. This would be trivial by using factorization theorem if the R-N derivative exits with respect to some sigma-finite measure.

40. Denote \( \nu \) as the counting measure restricted on \( \{0, 1, 2, \ldots\} \), and \( \delta(x) = I_{\{0,1\}}(x) \). Then
\[ \frac{dP_\theta^X}{d\nu}(x) = \left\{ e^{\theta_1 x_1} \right\}^{2-\theta_2} \cdot \left\{ \theta_1 (1 - \theta_1)^{1-x} \delta(x) \right\}^{\theta_2-1}, \text{a.e.}(\nu). \]

After some algebra,
\[ f_{\theta_1, \theta_2}(x_1, \ldots, x_n) = \frac{d(x^\frac{\theta_1}{\theta_2} P_\theta^X)}{d\nu^\frac{\theta_1}{\theta_2}}(x_1, \ldots, x_n) = \frac{e^{-\theta_1\theta_2} \cdot (1 - \theta_1)^{\theta_2-1}}{\prod_i x_i!} \cdot \theta_1 (1 - \theta_1)^{1-\theta_2} \sum_i x_i \cdot \prod_i \delta(x_i) x_i! \]
\[ \theta_2-1 \]
Thus \( (\sum_i X_i, \prod_i \delta(X_i)) = (\sum_i X_i, \prod_i \delta(X_i)) \) is sufficient by factorization theorem \((0! = 1! = 1 \text{ and } \delta(x) = I_{\{0,1\}}(x)) \). The minimal sufficient is straightforward using Theorem 55 by looking at the R-N derivative given above. Specifically, suppose \( f_{\theta_1, \theta_2}(x_1, \ldots, x_n)/f_{\theta_1, \theta_2}(y_1, \ldots, y_n) \) does not depend on \( (\theta_1, \theta_2) \in (0, 1) \times \{1, 2\} \). Let \( \theta_2 = 1 \), we get \( \sum_i x_i = \sum_i y_i \) and further obtain \( \Pi_i \delta(x_i) = \Pi_i \delta(y_i) \).

41. Denote \( P_\theta^X \) as the probability distribution of \( X \) here, then
\[ \frac{dP_\theta^X}{dm}(x) = \phi(x - \theta)^{I\{\theta \neq 0\}} \phi(x/\sqrt{2})^{I\{\theta = 0\}} = \phi \left( \frac{x - \theta}{1 + I\{\theta = 0\}} \right), \text{a.e.}(m), \]
where \( \phi \) is the standard normal density and \( m \) is the Lebesgue measure. Then the R-N derivative of \( x^\frac{\theta}{\theta_2} P_\theta^X \) with respect to \( m \) would be, after some algebra,
\[ \frac{d(x^\frac{\theta_1}{\theta_2} P_\theta^X)}{dm^\frac{\theta_1}{\theta_2}} = (2\pi)^{n/2} \exp \left\{ - \sum_i x_i^2 / 4 - n I\{\theta \neq 0\} \theta^2 / 2 \right\} \exp \left\{ -I\{\theta \neq 0\} \sum_i x_i^2 / 4 + I\{\theta \neq 0\} \theta \sum_i x_i \right\}, \text{a.e.}(m^n). \]
This implies that \( \{x^\frac{\theta_1}{\theta_2} P_\theta^X\}_{\theta \in \Theta} \) is an exponential family. However, \( \Gamma_\theta = \{I\{\theta \neq 0\}, I\{\theta \neq 0\} \theta \} \) does not contain any open rectangle in \( \mathbb{R}^2 \). Thus we can not utilize Claim 71. The good news is that Theorem 62 comes in handy. Consider \( \Theta' = \{\theta \neq 0 \mid \theta \in \mathbb{R} \} \), and denote \( \Theta = \mathbb{R} \). It is easy to see that \( P_\theta(B) = 0, \forall \theta \in \Theta' \Rightarrow P_\theta(B) = 0, \forall \theta \in \Theta \) (Normal distributions dominate each other). While for the smaller family \( \{x^\frac{\theta_1}{\theta_2} P_\theta^X\}_{\theta \in \Theta'} \), it can be seen that, for \( \theta \in \Theta' \),
\[ \frac{d(x^\frac{\theta_1}{\theta_2} P_\theta^X)}{dm^\frac{\theta_1}{\theta_2}} = (2\pi)^{n/2} \exp \left\{ - \sum_i x_i^2 / 2 - n \theta^2 / 2 \right\} \cdot \exp \left\{ \theta \sum_i x_i \right\}, \text{a.e.}(m^n). \]
which actually implies $\sum_i X_i$ is minimal sufficient and complete for the smaller family because $\Gamma_\theta = (-\infty, 0) \cup (0, \infty)$ contains open interval in $\Theta'$. By Theorem 62, $\sum_i X_i$ is complete for $\theta \in \Theta = \mathbb{R}$. It would not be surprising that $\sum_i X_i$ is not sufficient for the larger family when $\theta \in \Theta$ (Actually, $(\sum_i X_i, \sum_i X_i^2)$ is sufficient for $\theta \in \Theta$).

To show that $\sum_i X_i$ is not sufficient for $\theta \in \Theta$. One can either use factorization theorem or prove it using the first principle, i.e., the conditional distribution of $X_1, \ldots, X_n$ given $\bar{X}$ depends on $\theta$. To argue using factorization theorem, noticing that if $\sum_i X_i$ is sufficient for $\theta \in \Theta$, then there exists some $g_\theta$ and $h$ such that

$$
\frac{d(x_{i=1}^n P_{\theta}^X)}{dm^n}(x_1, \ldots, x_n) = g_\theta(\sum_i x_i) h(x_1, \ldots, x_n)
$$

which says $\sum_i x_i = \sum_i y_i \Rightarrow \frac{d(x_{i=1}^n P_{\theta}^X)}{dm^n}(x_1, \ldots, x_n) = \frac{d(x_{i=1}^n P_{\theta}^X)}{dm^n}(y_1, \ldots, y_n)$ does not depend on $\theta$. The ratio actually is, from the R-N derivative expression given above, $\exp\{1 + I\{\theta \neq 0\}\}(\sum_i y_i^2 - \sum_i x_i^2)$ which depends on $\theta$. Thus $\sum_i X_i$ is not sufficient for $\theta \in \Theta$.

Alternatively, As $X_1, \ldots, X_n$ are actually iid $N(\theta, 1+I\{\theta = 0\})$, the conditional distribution of $(X_1, \ldots, X_n)$ given $\bar{X}$ is still multivariate normal with mean $1\bar{X}$ and covariance $(1 + I\{\theta = 0\})(I_n - 11'/n)$, where $1$ is the $n$-dimensional vector with all elements being 1 and $I_n$ is $n \times n$ identity matrix. One could also see that the conditional distribution does not depend on $\theta$ when restricted on $\Theta'$. A good reference for this would be MVN facts from Dr. Vardeman’s Stat 542 page.

It is possible to prove the completeness of $\bar{X}$ directly from definition of completeness. That would end up with a similar proof process as proving the completeness of statistics for exponential family.

42. The likelihood function corresponding to $X_1, \ldots, X_n$ is $L(X; \mu, \sigma_1^2, \sigma_2^2) = \prod_{i=1}^n \phi\{(X_{1i} - \mu)/\sigma_1\} \phi\{(X_{2i} - \mu)/\sigma_2\}$, where $\phi$ is the standard normal density here and $X_i = (X_{1i}, X_{2i})', i = 1, \ldots, n$. By factorization theorem, we can see that $(\bar{X}_1, \bar{X}_2, S_1^2, S_2^2)'$ is sufficient (Or from Claim 69). From the fact that

$$
L(X; \mu, \sigma_1^2, \sigma_2^2) = \exp\left\{ \frac{\sum_i Y_{1i}^2 - \sum_i X_{1i}^2}{2\sigma_1^2} + \mu \frac{\sum_i X_{1i} - \sum_i Y_{1i}}{\sigma_1^2} + \frac{\sum_i Y_{2i}^2 - \sum_i X_{2i}^2}{2\sigma_2^2} + \mu \frac{\sum_i X_{2i} - \sum_i Y_{2i}}{\sigma_2^2} \right\} L(Y; \mu, \sigma_1^2, \sigma_2^2),
$$

utilizing Theorem 55, we can see that $T(X) = (\bar{X}_1, \bar{X}_2, S_1^2, S_2^2)'$ is minimal sufficient. To show that this statistic is not minimal sufficient, we just need to find some nontrivial bounded function of $T$ that is an unbiased estimator of 0. Let $f(a, b) = (a - b)I\{|a - b| \leq 1\}$, consider $h(x_1, x_2, x_3, x_4) = f((x_1 - x_2)/\sqrt{x_3 + x_4})$. $h$ is bounded and $E_\theta[h(T(X))] = 0$ for all $\theta = (\theta, \sigma_1^2, \sigma_2^2)$ because $h(T(X))$ is essentially the two sample t-test statistic, whose distribution is symmetric due to those two groups having the same population mean. However, $h \circ T$ is not 0. One candidate for the 1st order ancillary statistic would be $\bar{X}_1 - \bar{X}_2$.

43. From factorization theorem, after checking the likelihood function, $T(X) = (\sum_i X_i, \sum_i X_i^2)$ would be a sufficient statistic. Also consider the Likelihood function $L(X; \gamma)$, and $L(X; \gamma) = k(X,Y)L(Y; \gamma)$ where $k(X,Y) = \exp\{(\sum_i Y_i^2 - \sum_i X_i^2)/(2\gamma^2) + (\sum_i X_i - \sum_i Y_i)/\gamma\}$. $K(X,Y)$ independent of $\lambda$ implies that $T(X) = T(Y)$. Thus $T(X)$ is minimal sufficient.
Models and Measures of Statistical Information

46.

\[ I_X(\theta_0) = E_{\theta_0} \left\{ \frac{\partial \log f_0(T,S)}{\partial \theta} \right\}_{\theta=\theta_0}^2 \]

\[ = E_{\theta_0} \left\{ \frac{f_0'(T,S)}{f_0(T,S)} \right\}^2 \]

\[ \geq E_{\theta_0} \left[ E_{\theta_0} \left\{ \frac{f_0'(T,S)}{f_0(T,S)} \right\} \right]^2 \]

\[ = \int_T \left[ E_{\theta_0} \left\{ \frac{f_0'(T,S)}{f_0(T,S)} \right\} \right]^2 g_0(t) d\nu(t) \]

\[ = \int_T \left[ \int_S \frac{f_0'(t,s)}{f_0(t,s)} f_0(s|t) d\gamma(s) \right]^2 g_0(t) d\nu(t) \]

\[ = \int_T \left[ \int_S f_0'(t,s) f_0(t,s) g_0(t) d\gamma(s) \right]^2 g_0(t) d\nu(t) \]

\[ = \int_T g_0'(t)^2 \]

\[ = I_T(\theta_0) \]

47. Because \( \{P, Q\} \) are discrete measures, then we can consider counting measure \( \nu \) restricted on a countable set \( X \), and both \( P \) and \( Q \) are dominated by \( \nu \) with corresponding R-N derivatives \( p, q \). For statistic \( T = T(X) \), let \( T \) be its corresponding area. Then

\[ I_X(P, Q) = \int_X p(x) \log \frac{p(x)}{q(x)} d\nu(x) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \]

\[ = \sum_{t \in T} \sum_{x: T(x)=t} p(x) \log \frac{p(x)}{q(x)} = \sum_{t \in T} \left\{ \sum_{x: T(x)=t} p(x) \sum_{x: T(x)=t} \frac{p(x)}{q(x)} \left( -\log \frac{q(x)}{p(x)} \right) \right\} \]

\[ \geq \sum_{t \in T} \left\{ \sum_{x: T(x)=t} p(x) \log \frac{\sum_{x: T(x)=t} p(x)}{\sum_{x: T(x)=t} q(x)} \right\} \]

\[ = I_T(X)(P^T, Q^T). \]

48. Consider \( \{P_0, P_0\} \) and let \( P_{0_1} = P, P_{0_2} = Q \). Both are dominated by \( \mu = \nu \times \gamma \) with R-N derivatives \( f_{0_1}, f_{0_2} \). Then

\[ I_X(P, Q) = E_{\theta_1} \log \frac{f_{\theta_1}(T,S)}{f_{\theta_2}(T,S)} \]

\[ = E_{\theta_1} \left\{ -\log \frac{f_{\theta_2}(T,S)}{f_{\theta_1}(T,S)} \right\} \]

\[ \geq E_{\theta_1} \left\{ -\log \left( \int_S \frac{f_{\theta_2}(T,s)}{f_{\theta_1}(T,S)} ds \right) \right\} \]

\[ = E_{\theta_1} \left\{ -\log \left( \int_S \frac{f_{\theta_2}(T,s)}{f_{\theta_1}(T,S)} g_{\theta_1}(T) ds \right) \right\} \]

\[ = E_{\theta_1} \log \frac{g_{\theta_2}(T)}{g_{\theta_1}(T)} \]

\[ = I_T(X)(P^T, Q^T). \]
\[ g_\beta(x) = \min(x, 1) f(x|\beta) I\{x > 0\} + I\{x = 0\} \int (1 - \min(y, 1)) f(y|\beta) dy, \]

where \( f(x|\beta) = \exp\{-x/\beta\}/\beta I\{x > 0\} \). Denote \( h(\beta) = \int (1 - \min(y, 1)) f(y|\beta) dy = 1 - \beta + \beta e^{-1/\beta} \), now the log of the density becomes

\[ \log g_\beta(x) = I\{x > 0\} \{\log(\min(x, 1)) + \log f(x|\beta)\} + I\{x = 0\} \log h(\beta), \]

then

\[ E_{\beta_0} \left\{ \left( \frac{\partial \log g_\beta(x)}{\partial \beta} \right)^2 \bigg|_{\beta = \beta_0} \right\} = \int \left( \frac{\partial \log g_\beta(x)}{\partial \beta} \right)^2 \bigg|_{\beta = \beta_0} g_{\beta_0}(x) d(\delta_0 + \lambda). \]

More specifically, it is not hard to see that the result has the form

\[ I_Y(\beta_0) + \int (-1 + \min(x, 1)) f(x|\beta_0)(\partial \log f(x|\beta)/\partial \beta|_{\beta = \beta_0})^2 dx + (\partial \log h(\beta)/\partial \beta)^2|_{\beta = \beta_0} h(\beta_0), \]

where

\[ I_Y(\beta_0) = 1/\beta_0^2 \]

\[ (\partial \log h(\beta)/\partial \beta)^2|_{\beta = \beta_0} h(\beta_0) = \frac{e^{-2/\beta_0}(\beta e^{1/\beta_0} - \beta_0 - 1)^2}{(1 - \beta_0 + \beta_0 e^{-1/\beta_0})\beta_0^2} \]

\[ \int (-1 + \min(x, 1)) f(x|\beta_0)(\partial \log f(x|\beta)/\partial \beta|_{\beta = \beta_0})^2 dx = \frac{1 + 2\beta_0 + 3\beta_0^2}{\beta_0^3} e^{-1/\beta_0} - \frac{3\beta_0 - 1}{\beta_0^2} \]

Intuitively, if \( P(B) = 1 \), then \( X \) and \( Y \) are the same and should hold the same fisher information matrix, that means \( e^{-1/\beta_0} = P(Y \geq 1) \approx 1 \), which impiles that \( \beta_0 \to \infty \). It can also be derived directly from the above results.

50. \( \theta = (p_Y, p_Z) \). \( \log f_\theta(y, z) = y \log p_Y + (n - y) \log(1 - p_Y) + z \log p_Z + (y - z) \log(1 - p_Z) + \log \left( \begin{array}{c} y \\ n \end{array} \right) + \log \left( \begin{array}{c} n - y \\ n \end{array} \right) \). One sees that \( \partial^2 \log f_\theta(y, z)/\partial p_Y \partial p_Z = \partial^2 \log f_\theta(y, z)/\partial p_Z \partial p_Y = 0 \) and \( \partial^2 \log f_\theta(y, z)/\partial p_Y^2 = -(n - y)(1 - p_Y)^2 \) and \( \partial^2 \log f_\theta(y, z)/\partial p_Z^2 = -(n - z)(1 - p_Z)^2 \). Notice further that \( E(Y) = np_Y, E(Z) = E\{E(Z|Y)\} = np_Zp_Y \). Denote \( \theta_0 = (p_Y^*, p_Z^*) \), then

\[ I_X(\theta_0) = E_{\theta_0} \left\{ -\frac{\partial^2}{\partial \theta \partial \theta'} \log f_\theta(X) \bigg|_{\theta = \theta_0} \right\} = \begin{pmatrix} \frac{n}{p_Y(1-p_Y)} & 0 \\ 0 & \frac{n}{p_Z(1-p_Z)} \end{pmatrix}. \]