PRINCIPLE OF INCLUSION-EXCLUSION: Let $A_1, \dots, A_n \in L$. Then it holds

$$P(\overset{\text{n}}{\underset{i=1}{\cup}} A_i) = \sum_{i=1}^{n} (-1)^{i-1} P_i,$$

where

$$\begin{aligned} P_1 &= \sum_{i=1}^{n} P(A_i); \\ P_2 &= \sum_{i_1<i_2}^{n} P(A_{i_1} \cap A_{i_2}) \\ P_3 &= \sum_{i_1<i_2<i_3}^{n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ &\vdots \\ P_n &= P(A_1 \cap \dots \cap A_n). \square \end{aligned}$$

BONFERRONI'S INEQUALITY: Let $A_1, \dots, A_n \in L$. Then it holds

$$\sum_{i=1}^{n} P(A_i) - \sum_{i_1<i_2}^{n} P(A_{i_1} \cap A_{i_2}) \leq P(\overset{\text{n}}{\underset{i=1}{\cup}} A_i) \leq \sum_{i=1}^{n} P(A_i).$$

Proof: ($\Longrightarrow$) We need to show

$$P(\overset{\text{n}}{\underset{i=1}{\cup}} A_i) \leq \sum_{i=1}^{n} P(A_i). \tag{1.1}$$

The case when $n = 1$ is trivial. Let $n \in \mathbb{N}$, $n > 1$, be arbitrary. From the Principle of Inclusion-Exclusion (PIE), it is

$$\begin{aligned} P(\overset{\text{n}}{\underset{i=1}{\cup}}) &= \sum_{i=1}^{n} (-1)^{n-1} P_i \\ &= \begin{cases} P_1 - (P_2 - P_3) - \dots - (P_{n-1} - P_n), & \text{if n is odd} \\ P_1 - (P_2 - P_3) - \dots - (P_{n-2} - P_{n-1}) - P_n, & \text{if n is even.} \end{cases} \end{aligned} \tag{1.2}$$

But, for $i > 1$, it holds

$$\begin{aligned}
P_i &= \sum_{j_1<j_2<\ldots<j_i}^{n} P(\overset{\text{i}}{\underset{k=1}{\cap}} A_{j_k}) \\
&= \sum_{j_1<j_2<\ldots<j_i}^{n} \left(\sum_{m>j_i}^{n} P(\overset{\text{i}}{\underset{k=1}{\cap}} A_{j_k} \cap A_m) + \sum_{m>j_i}^{n} P(\overset{\text{i}}{\underset{k=1}{\cap}} A_{j_k} \cap A_m^c) \right) \\
&= P_{i+1} + \delta \\
&\geq P_{i+1}, \qquad (1.3)
\end{aligned}$$

where $\delta \geq 0$. The result of (1.1) immediately follows from (1.2) and (1.3), since

$$-\sum_{i=2}^{[\frac{n}{2}]+1} (P_i - P_{i+1}) - 1_{\{n \text{ is even}\}} \cdot P_n \leq 0,$$

where $[\![n]\!]$ is the greatest integer contained in n.

($\Longleftarrow$) We need to show

$$\sum_{i=1}^{n} P(A_i) - \sum_{i_1<i_2}^{n} P(A_{i_1} \cap A_{i_2}) \leq P(\overset{\text{n}}{\underset{i=1}{\cup}} A_i). \qquad (1.4)$$

The result immediately follows for $n = 1$, since $P_2 = \emptyset$, which implies that $P(A_1) \leq P(A_1)$. Suppose that $n \in \mathbb{N}$ is arbitrary, $n > 1$. From the PIE, it holds

$$\begin{aligned}
P(\overset{\text{n}}{\underset{i=1}{\cup}}) &= \sum_{i=1}^{n} (-1)^{n-1} P_i \\
&= \begin{cases} P_1 - P_2 + (P_3 - P_4) + \ldots + (P_{n-1} - P_n)\,, & \text{if n is even} \\ P_1 - P_2 + (P_3 - P_4) + \ldots + (P_{n-2} - P_{n-1}) + P_n, & \text{if n is odd.} \end{cases} \qquad (1.5)
\end{aligned}$$

The result of (1.4) immediately follows from (1.3) and (1.5), since

$$\sum_{i=3}^{[\frac{n}{2}]+1} (P_i - P_{i+1}) + 1_{\{n \text{ is odd}\}} \cdot P_n \geq 0. \square$$

NEYMAN-PEARSON LEMMA: Suppose we wish to test $H_0 : \boldsymbol{X} \sim f_{\theta_0}(\boldsymbol{x})$, versus $H_1 : \boldsymbol{X} \sim f_{\theta_1}(\boldsymbol{x})$, where f_{θ_i} is the pdf (or pmf) for $\boldsymbol{X}$ under H_i, $i = 0, 1$, where both H_0 and H_1 are each simple.

(i) Any test of the form

$$\phi(\boldsymbol{x}) = \begin{cases} 1, \text{if} \, f_{\theta_1}(\boldsymbol{x}) > k \cdot f_{\theta_0}(\boldsymbol{x}) \\ \gamma(\boldsymbol{x}), \text{if} \, f_{\theta_1}(\boldsymbol{x}) = k \cdot f_{\theta_0}(\boldsymbol{x}) \\ 0, \text{if} \, f_{\theta_1}(\boldsymbol{x}) < k \cdot f_{\theta_0}(\boldsymbol{x}) \end{cases} \tag{1.6}$$

for some $k \geq 0$, and $0 \leq \gamma(\boldsymbol{x}) \leq 1$, is most powerful of its significance level for testing H_0 versus H_1. If $k = \infty$, the test

$$\phi(\boldsymbol{x}) = \begin{cases} 1, \text{if} \, f_{\theta_0}(\boldsymbol{x}) = 0 \\ 0, \text{if} \, f_{\theta_0}(\boldsymbol{x}) > 0 \end{cases} \tag{1.7}$$

is most powerful of size (or significance level) 0 for testing H_0 versus H_1.

(ii) Given $0 \leq \alpha \leq 1$, there exists a test of form (1.6) or (1.7), with $\gamma(\boldsymbol{x}) = \gamma$ (i.e., a constant) such that

$$E_{\theta_0}(\phi(\boldsymbol{X})) = \alpha.$$

Chapter 1

KEY THEOREMS: PART 2

BINOMIAL THEOREM: If $n \in \mathbb{N} \cup \{0\}$, then for a arbitrary real number x, it holds

$$(1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i. \tag{1.1}$$

Proof: Here, let x be an arbitrary real number and let $n \in \mathbb{N} \cup \{0\}$ be arbitrary. Let $f(x) = (1+x)^n$. Then, the Maclaurin series for f is given by

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(0)(x-0)^i}{i!}, \tag{1.2}$$

where $f^{(0)}(0) = f(0)$. It is,

$$\begin{aligned} f'(x) &= n(1+x)^{n-1} \\ f''(x) &= n(n-1)(1+x)^{n-2} \\ &\vdots \\ f^{(k)}(x) &= n(n-1)\cdots(n-k+1)(1+x)^{n-k} \cdot 1_{\{k \le n\}}. \end{aligned}$$

Thus, re-writing (1.2), it holds

$$\begin{aligned} f(x) &= \sum_{i=0}^{\infty} \frac{f^{(i)}(0)(x-0)^i}{i!} \\ &= f(0) + \sum_{i=1}^{n} \frac{n(n-1)\cdots(n-i+1)(1+0)^{n-i}x^i}{i!} \\ &= \sum_{i=0}^{n} \frac{n!x^i}{(n-i)!i!} \\ &= \sum_{i=0}^{n} \binom{n}{i} x^i, \end{aligned}$$

as desired. Since x and n were chosen arbitrarily, this result holds for all $x \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$. Therefore, if $x \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$, (1.1) holds.□

MULTIPLICATION RULE: Suppose $A_1, \ldots, A_n \in L$ and $P(A_1 \cap \cdots \cap A_{n-1}) > 0$. Then, it holds

$$P(\bigcap_{i=1}^{\text{n}} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}). \tag{1.3}$$

Proof: By induction *w.r.t.* n. To establish the basis for induction, suppose $n = 2$. It is,

$$P(A_1 \cap A_2) = \frac{P(A_1 \cap A_2)}{P(A_1)} P(A_1) \stackrel{*}{=} P(A_2|A_1)P(A_1),$$

where ($\stackrel{*}{=}$) holds by the definition of conditional probability. This establishes the basis for induction. Next, suppose that the result (expression (1.3)) holds for some $n \in \mathbb{N}$, $n > 2$. We need to show that (1.3) holds for $(n+1) \in \mathbb{N}$. Let $B = \bigcap_{i=1}^{\text{n}} A_i$. It is,

$$P(\bigcap_{i=1}^{\text{n+1}} A_i) = P(B \cap A_{n+1}) = \frac{P(A_{n+1} \cap B)}{P(B)} P(B) \stackrel{*}{=} P(A_{n+1}|B)P(B), \tag{1.4}$$

where ($\stackrel{*}{=}$) holds by the definition of conditional probability. Also, by the induction hypothesis, it holds

$$P(B) = P(\bigcap_{i=1}^{\text{n}} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}),$$

so that (1.4) becomes

$$P(\bigcap_{i=1}^{\text{n+1}} A_i) = P(B)P(A_{n+1}|B) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots P(A_{n+1}|A_1 \cap \cdots \cap A_n),$$

which establishes the induction step. Therefore, by mathematical induction, for all $n \in \mathbb{N}$, $n \geq 2$, if $A_1, ..., A_n \in L$ and $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, then

$$P(\overset{\text{n}}{\underset{i=1}{\cap}} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}).\square$$

LAW OF TOTAL PROBABILITY: Let $\{H_i\}_{i=1}^{\infty}$ be a partition of Ω, for which $P(H_i) > 0$ for all $i \in \mathbb{N}$. If $A \in L$, it holds

$$P(A) = \sum_{i=1}^{\infty} P(A \cap H_i) = \sum_{i=1}^{\infty} P(A|H_i)P(H_i). \tag{1.5}$$

Proof: For each $i \in \mathbb{N}$, since $P(H_i) > 0$, it is

$$P(A \cap H_i) = \frac{P(A \cap H_i)}{P(H_i)} P(H_i) \overset{*}{=} P(A|H_i)P(H_i),$$

where ($\overset{*}{=}$) holds by the definition of conditional probability. Thus, it is clear that the two summands of (1.5) are equal. Hence, it is left to show that

$$P(A) = \sum_{i=1}^{\infty} P(A \cap H_i).$$

It is,

$$A = (A \cap H_1) \cup (A \cap H_2) \cup \cdots \cup (A \cap H_i) \cup \cdots = \overset{\infty}{\underset{i=1}{\cup}} (A \cap H_i),$$

and by PIE, it holds

$$P(A) = P(\overset{\infty}{\underset{i=1}{\cup}} (A \cap H_i)) = \sum_{j=1}^{\infty} (-1)^{j-1} P_j, \tag{1.6}$$

where

$$P_1 = \sum_{i=1}^{\infty} P(A \cap H_i);$$

$$P_2 = \sum_{i_1 < i_2}^{\infty} P(A \cap H_{i_2} \cap A \cap H_{i_2});$$

$$\vdots$$

$$P_j = \sum_{i_1 < i_2 < \ldots < i_j}^{\infty} P(A \cap H_{i_2} \cap A \cap H_{i_2} \cap \cdots \cap A \cap H_{i_j});$$

$$\vdots$$

But, $P_j = 0$, for all $j > 1$, since the H_i are disjoint. The result follows immediately. Therefore, if $\{H_i\}_{i=1}^{\infty}$ is a partition of Ω, for which $P(H_i) > 0$ for all $i \in \mathbb{N}$ and $A \in L$, it holds

$$P(A) = \sum_{i=1}^{\infty} P(A \cap H_i) = \sum_{i=1}^{\infty} P(A|H_i)P(H_i). \square$$

BAYES' RULE: Let $\{H_i\}_{i=1}^{\infty}$ be a partition of Ω, such that $P(H_i) > 0$ for all $i \in \mathbb{N}$, and let $A \in L$ for which $P(A) > 0$. Then, for all $i \in \mathbb{N}$, it holds

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum\limits_{i=1}^{\infty} P(A|H_i)P(H_i)}.$$

Proof: Since $P(A) > 0$, by the definition of conditional probability, it follows

$$P(H_i|A) = \frac{P(H_i \cap A)}{P(A)} \overset{*}{=} \frac{P(A|H_i)P(H_i)}{\sum\limits_{i=1}^{\infty} P(A|H_i)P(H_i)},$$

where the numerator of $(\overset{*}{=})$ holds by the definition of conditional probability and the denomiator of $(\overset{*}{=})$ holds by the law of total probability. Therefore, if $\{H_i\}_{i=1}^{\infty}$ is a partition of Ω, such that $P(H_i) > 0$ for all $i \in \mathbb{N}$, and $A \in L$ for which $P(A) > 0$, then

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum\limits_{i=1}^{\infty} P(A|H_i)P(H_i)}.$$

$\blacksquare$

Chapter 1

KEY THEOREMS: PART 3

MARKOV'S INEQUALITY: Let $h(X)$ be a non-negative Borel-measurable function of a random variable X. If $E(h(X))$ exists, then it holds:

$$P(h(X) \geq \epsilon) \leq \frac{E(h(X))}{\epsilon} \quad \forall \epsilon > 0. \tag{1.1}$$

When $h(X) = |X|^r$ and $\epsilon = k^r$, where $r, k > 0$, then (1.1) reduces to

$$P(|X| \geq k) \leq \frac{E(|X|^r)}{k^r},$$

Markov's Inequality.

Proof (continuous case): Let χ be the support of X. Since h is a non-negative function, define the set $A \subseteq \chi$ by

$$A = \{x \in \chi : h(x) \geq \epsilon\}.$$

Also, let $A^c = \chi \backslash A$, the complement of A. Suppose $E(g(X))$ exists. It is,

$$\begin{aligned}
E(h(X)) &= \int_A h(x) f_X(x) dx + \int_{A^c} h(x) f_X(x) dx \\
&\overset{(A)}{\geq} \int_A h(x) f_X(x) dx \\
&\geq \int_A \epsilon \cdot f_X(x) dx \\
&= \epsilon \cdot P(h(X) \geq \epsilon),
\end{aligned} \tag{1.2}$$

where (A) holds since h is non-negative. Dividing both sides of the inequality (1.2) by ϵ, yields (1.1). When $h(X) = |X|^r$ and $\epsilon = k^r$, where $r, k > 0$, we have

$$P(|X| \geq k) = P(|X|^r \geq k^r) \leq \frac{E(|X|^r)}{k^r}, \tag{1.3}$$

as required. Therefore, if h is a non-negative Borel-measurable function, (1.1) holds. Also, when $h(X) = |X|^r$ and $\epsilon = k^r$, where $r, k > 0$, it follows that (1.3) (Markov's Inequality) holds.□

CHEBYCHEV'S INEQUALITY: Let $h(X) = (X - \mu)^2$ and $\epsilon = k^2\sigma^2$, where $E(X) = \mu$, $\text{Var}(X) = \sigma^2 < \infty$, and $k > 0$. Then it holds

$$P(|X - \mu| \geq k \cdot \sigma) \leq \frac{1}{k^2}. \tag{1.4}$$

Proof: From Markov's Inequality, it is

$$\begin{aligned} P(|X - \mu| \geq k \cdot \sigma) &= P(|X - \mu|^2 \geq k^2\sigma^2) \\ &\leq \frac{E((X - \mu)^2)}{k^2\sigma^2} \\ &= \frac{\text{Var}(X)}{k^2\sigma^2} \\ &= \frac{1}{k^2}, \end{aligned}$$

as required. Therefore, if $\text{Var}(X)$ exists, (1.4) holds.□

LYAPUNOV'S INEQUALITY: Let $0 < E(|X|^n) < \infty$. For an arbitrary $k \in \mathbb{N}$, $2 \leq k \leq n$, it holds

$$\left(E(|X|^{k-1})\right)^{\frac{1}{k-1}} \leq \left(E(|X|^k)\right)^{\frac{1}{k}}.$$

Proof: SEE STAT 6710 COURSE NOTES.

LEMMA 4.8.1: Let a and b be positive numbers and $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then it holds that

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq a \cdot b,$$

with equality holding iff $a^p = b^q$.

Proof: Fix b, and let $g(a) = p^{-1}a^p + q^{-1}b^q - a \cdot b$. We proceed by minimizing g. Here,

$$g'(a) = a^{p-1} - b = 0 \quad \iff \quad a^{p-1} = b. \tag{1.5}$$

Raising each side of the equation – given by (1.5) – to the q^{th} power, we have

$$a^{p-1} = b \quad \iff \quad a^{q \cdot p - q} = b^q \quad \iff \quad a^p = b^q, \tag{1.6}$$

since $p \cdot q = p + q$. Differentiating (1.5), it is

$$g''(a) = (p-1)\, a^{p-2} > 0,$$

since $a > 0$ and $p, q > 1$. This implies (along with (1.5) and (1.6)) that $a^p = b^q$ is the unique minimum for g. So,

$$\begin{aligned} g(a^{p-1} = b) &= \frac{1}{p} a^p + \frac{1}{q} \left(a^{p-1}\right)^q - a \cdot a^{p-1} \\ &= \frac{1}{p} a^p + \frac{1}{q} a^p - a^p \\ &= a^p \left(\frac{1}{p} + \frac{1}{q} - 1 \right) \\ &= 0. \end{aligned}$$

Thus, $g(a) \geq 0$. This implies that

$$\frac{1}{p} a^p + \frac{1}{q} b^q - a \cdot b \geq 0 \qquad \Longleftrightarrow \qquad \frac{1}{p} a^p + \frac{1}{q} b^q \geq a \cdot b,$$

as desired.□

HÖLDER'S INEQUALITY: Let X, Y be two random variables. Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then, it holds that

$$E(|X \cdot Y|) \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}.$$

Proof: Define a and b as follows:

$$a = \frac{|X|}{(E(|X|^p))^{1/p}} > 0; \quad b = \frac{|Y|}{(E(|Y|^q))^{1/q}} > 0.$$

From Lemma 4.8.1, it is

$$\begin{aligned} & \frac{1}{p} \frac{|X|^p}{E(|X|^p)} + \frac{1}{q} \frac{|Y|^q}{E(|Y|^q)} \geq \frac{|X|}{(E(|X|^p))^{1/p}} \cdot \frac{|Y|}{(E(|Y|^q))^{1/q}} \\ \Longleftrightarrow \quad & \frac{1}{p} + \frac{1}{q} \geq \frac{E(|X \cdot Y|)}{(E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}} \\ \Longleftrightarrow \quad & E(|X \cdot Y|) \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}, \end{aligned}$$

since $\frac{1}{p} + \frac{1}{q} = 1$. This establishes the desired result.□

MINKOWSKI'S INEQUALITY: Let X, Y be two random variables. Then, for $1 \leq p < \infty$, it holds

$$(E(|X+Y|^p))^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}.$$

Proof: Suppose $p = 1$. From the triangle inequality, it holds

$$|X+Y| \leq |X| + |Y| \qquad \Longleftrightarrow \qquad E(|X+Y|) \leq E(|X|) + E(|Y|),$$

which establishes the desired result. Next, suppose that $p > 1$, and let $q = \frac{p}{p-1}$. This implies that $\frac{1}{q} + \frac{1}{p} = 1$. From the triangle inequality, it holds

$$|X+Y|^p = |X+Y| \cdot |X+Y|^{p-1} \leq (|X| + |Y|)\,|X+Y|^{p-1}.$$

Thus, from Hölder's Inequality, we have

$$\begin{aligned} E(|X+Y|^p) &\leq E(|X||X+Y|^{p-1}) + E(|Y||X+Y|^{p-1}) \\ &\leq (E(|X|^p))^{1/p}\Big(E(|X+Y|)^{q(p-1)}\Big)^{1/q} + (E(|Y|^p))^{1/p}\Big(E(|X+Y|)^{q(p-1)}\Big)^{1/q} \\ &= (E(|X|^p))^{1/p}(E(|X+Y|)^p)^{1/q} + (E(|Y|^p))^{1/p}(E(|X+Y|)^p)^{1/q}. \end{aligned}$$

Dividng both sides of this inequality by $(E(|X+Y|)^p)^{1/q}$, we have

$$\frac{E(|X+Y|^p)}{E(|X+Y|^p)^{1/q}} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p},$$

which – since $1 - \frac{1}{q} = \frac{1}{p}$ – provides that

$$E(|X+Y|^p)^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p},$$

as required.□

CAUCHY-SCHWARTZ-INEQUALITY: Let X, Y be two random variables with finite variance. Then it holds:

(i) $\mathrm{Cov}(X, Y)$ exists.

(ii) $(E(X \cdot Y))^2 \leq E(X^2)E(Y^2)$.

(iii) $(E(X \cdot Y))^2 = E(X^2)E(Y^2)$ iff there exists an $(\alpha, \beta) \in \mathbb{R}^2 \backslash \{(0,0)\}$, such that $P(\alpha X + \beta Y = 0) = 1$.

Proof: (i) Since X and Y each have finite variances, it follows that for $i = 1, 2$, $E(X^i) < \infty$ and $E(Y^i) < \infty$. Thus, from Hölder's Inequality, taking $p = q = 2$, it holds

$$E(|X \cdot Y|) \leq \left(E(|X|^2)\right)^{1/2} \left(E(|Y|^2)\right)^{1/2} < \infty, \tag{1.7}$$

so that $E(X \cdot Y)$ exists. Hence,

$$E(X \cdot Y) - E(X)E(Y) < \infty,$$

so that $\mathrm{Cov}(X, Y)$ exists.
(ii) From (1.7), it is,

$$(E(|X \cdot Y|))^2 \leq E(|X|^2)E(|Y|^2) = E(X^2)E(Y^2). \tag{1.8}$$

Since the absolute value function is convex, Jensen's Inequality provides that

$$|E(X \cdot Y)| \leq E(|X \cdot Y|) \qquad \Longleftrightarrow \qquad (E(X \cdot Y))^2 \leq (E(|X \cdot Y|))^2$$

Therefore, from (1.8), it holds

$$(E(X \cdot Y))^2 \leq (E(|X \cdot Y|))^2 \leq E(X^2)E(Y^2),$$

as desired.
(iii) From the course notes for (ii) (note that the proof given above is not that given in class), a necessary and sufficient condition for equality in (ii) is that $E((\alpha \cdot X + \beta \cdot Y)^2) = 0$. Thus, it suffices to show that $E((\alpha \cdot X + \beta \cdot Y)^2) = 0$ is equivalent to $P(\alpha \cdot X + \beta \cdot Y = 0) = 1$.
$(\Longrightarrow)$ Let $Z = \alpha \cdot X + \beta \cdot Y$. Since $0 = E((\alpha \cdot X + \beta \cdot Y)^2) = E(Z^2) = \mathrm{Var}(Z) + (E(Z))^2$, and $\mathrm{Var}(Z), (E(Z))^2 \geq 0$, it follows that $\mathrm{Var}(Z) = E(Z) = 0$. Thus, $Z \sim \mathrm{Dirac}(0)$. Hence,

$$1 = P(Z = 0) = P(\alpha \cdot X + \beta \cdot Y = 0),$$

as desired.
$(\Longleftarrow)$ Suppose that $P(\alpha \cdot X + \beta \cdot Y = 0) = 1$, for $(\alpha, \beta) \in \mathbb{R} \backslash \{(0, 0)\}$. Then, $P(X = -\frac{\beta \cdot Y}{\alpha}) = 1$, which implies

$$\begin{aligned}
(E(X \cdot Y))^2 &= \left(E(-\frac{\beta \cdot Y}{\alpha} \cdot Y)\right)^2 \\
&= \left(\frac{\beta}{\alpha}\right)^2 \left(E(Y^2)\right)^2 \\
&= \left(\frac{\beta}{\alpha}\right)^2 E(Y^2)E(Y^2) \\
&= E(X^2)E(Y^2),
\end{aligned}$$

as desired. Therefore, (iii) holds.$\square$

SLUTSKY'S THEOREM: Let $\{X_n\}_{n=1}^{\infty}$ and $\{Y_n\}_{n=1}^{\infty}$ be sequences of random variables, and X be a random variable, all of which are defined on a probability space (Ω, L, P). Let $c \in \mathbb{R}$ be a constant. Then, it holds:

(i) $X_n \xrightarrow{d} X,\ Y_n \xrightarrow{p} c \implies X_n + Y_n \xrightarrow{d} X + c.$

(ii) $X_n \xrightarrow{d} X,\ Y_n \xrightarrow{p} c \implies X_n Y_n \xrightarrow{d} c \cdot X.$ If $c = 0$, then also $X_n Y_n \xrightarrow{p} 0.$

(iii) $X_n \xrightarrow{d} X,\ Y_n \xrightarrow{p} c \implies \frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$, provided that $c \neq 0$.

Proof: (i) Since $X_n \xrightarrow{d} X$, it follows (via Theorem result) that $X_n + c \xrightarrow{d} X + c$. Also, $Y_n \xrightarrow{p} c$, implies (via Theorem result) that $Y_n - c \xrightarrow{p} 0$. It holds,

$$Y_n - c = Y_n + X_n - (X_n + c) \xrightarrow{p} 0,$$

so that Theorem 6.1.17 of the course notes provides that $Y_n + X_n \xrightarrow{d} X + c$. This is what we needed to show.

(ii) Suppose $c \neq 0$. By Theorem 6.1.11(x), since $Y_n \xrightarrow{p} c$, it follows that $Y_n X_n \xrightarrow{p} c \cdot X_n$. Moreover, by Theorem 6.1.9(ii), it follows that $c \cdot X_n \xrightarrow{d} c \cdot X$. It is,

$$X_n Y_n - c \cdot X_n = (X_n Y_n - c \cdot X) - (c \cdot X_n - c \cdot X) \xrightarrow{p} 0,$$

and since $c \cdot X_n \xrightarrow{d} c \cdot X$, Theorem 6.1.17 provides that $X_n Y_n \xrightarrow{d} c \cdot X$, as desired. Next, suppose that $c = 0$. Let $\epsilon, k > 0$ be arbitrary real numbers. It follows that

$$\begin{aligned} P(|X_n Y_n| > \epsilon) &= P(|X_n Y_n| > \epsilon, Y_n \leq \frac{\epsilon}{k}) + P(|X_n Y_n| > \epsilon, Y_n > \frac{\epsilon}{k}) \\ &\leq P(|X_n \frac{\epsilon}{k}| > \epsilon) + P(|Y_n| > \frac{\epsilon}{k}) \\ &\leq P(|X_n| > k) + P(|Y_n| > \frac{\epsilon}{k}). \end{aligned}$$

Since $Y_n \xrightarrow{p} 0$ and $X_n \xrightarrow{d} X$, then for any fixed $k > 0$, it holds

$$\lim_{n \to \infty} P(|X_n Y_n| > \epsilon) \leq \lim_{n \to \infty} P(|X_n| > k).$$

But, we can make $P(|X_n| > k)$ as small as we like, by choosing k large. Thus,

$$\lim_{n \to \infty} P(|X_n Y_n| > \epsilon) \leq \lim_{n \to \infty} P(|X_n| > k) = 0,$$

which provides that $X_n Y_n \xrightarrow{p} 0$.

(iii) Suppose $c \neq 0$. Let $Z_n \xrightarrow{p} 1$. Let $Y_n = c \cdot Z_n$. Then, $\frac{1}{Y_n} = \frac{1}{Z_n} \cdot \frac{1}{c}$. By parts (v) and (viii) of Theorem 6.1.11, it follows that $\frac{1}{Y_n} \xrightarrow{p} \frac{1}{c}$. From part (ii) above, with $X_n \xrightarrow{d} X$ and $\frac{1}{Y_n} \xrightarrow{p} \frac{1}{c}$, implies that $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$.□

WEAK LAW OF LARGE NUMBERS (VERSION 1): Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of *i.i.d.* random variables with mean $E(X_i) = \mu$ and variance $\text{Var}(X_i) = \sigma^2 < \infty$. Let $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then, it holds

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0 \quad \forall \epsilon > 0. \tag{1.9}$$

That is, $\overline{X}_n$ is consistent for μ.

Proof: Note that since the X_i are *i.i.d.*, it holds that $E(\overline{X}_n) = \mu$ and $\text{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$. Let $\epsilon > 0$ be arbitrary. From Markov's Inequality, we have

$$\begin{aligned} P(|\overline{X}_n - \mu| \geq \epsilon) &\leq \frac{E\big(\big(|\overline{X}_n - \mu|\big)^2\big)}{\epsilon^2} \\ &= \frac{\text{Var}(\overline{X}_n)}{\epsilon^2} \\ &= \frac{\sigma^2}{n \cdot \epsilon^2}. \end{aligned}$$

Thus,

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| \geq \epsilon) \leq \lim_{n \to \infty} \frac{\sigma^2}{n \cdot \epsilon^2} = 0,$$

as desired. Since $\epsilon > 0$ was chosen arbitrarily, (1.9) holds.□

KHINTCHINE'S WLLN: Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of *i.i.d.* random variables with finite mean, $E(X_i) = \mu$. Then it holds:

$$\overline{X}_n = \frac{1}{n} T_n \xrightarrow{p} \mu. \square$$

KOLMOGOROV'S STRONG LAW OF LARGE NUMBERS: Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of *i.i.d.* random variables. Let $T_n = \sum_{i=1}^{n} X_i$. Then, it holds

$$\frac{T_n}{n} = \overline{X}_n \xrightarrow{a.s.} \mu < \infty \qquad \Longleftrightarrow \qquad E(|X|) < \infty \quad (\text{ and then } \mu = E(X)). \square$$

LINDEBERG-LÉVY CENTRAL LIMIT THEOREM: Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of *i.i.d.* random variables with $E(X_i) = \mu$ and $0 < \text{Var}(X_i) = \sigma^2 < \infty$. Then, it holds for $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ that

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \xrightarrow{d} Z,$$

where $Z \sim N(0,1)$.$\square$

FACTORIZATION CRITERION: Let $X_1, ..., X_n$ be random variables with pdf (or pmf) $f(x_1, ..., x_n|\theta)$, $\theta \in \Theta$. Then, $T(X_1, ..., X_n)$ is sufficient for θ iff we can write

$$f(x_1, ..., x_n|\theta) = h(x_1, ..., x_n)g(T(x_1, ..., x_n)|\theta),$$

where h does not depend on θ and g does not depend on $x_1, ..., x_n$, except as a function of T.$\square$

RAO-BLACKWELL: Let $\{F_\theta : \theta \in \Theta\}$ be a family of cdf's, and let h be any statistic in U, where U is the non-empty class of unbiased estimators of θ with $E_\theta(h^2) < \infty$. Let T be a sufficient statistic for $\{F_\theta : \theta \in \Theta\}$. Then, the conditional expectation, $E_\theta(h|T)$, is independent of θ and it is an unbiased estimate of θ. Additionally,

$$E_\theta((E(h|T) - \theta)^2) \leq E_\theta((h - \theta)^2) \quad \forall \theta \in \Theta,$$

with equality holding iff $h = E(h|T)$.$\square$

LEHMANN-SCHEFFÉE: If T is a complete sufficient statistic and if there exists an unbiased estimate h of θ, then $E(h|T)$ is the (unique) UMVUE.

Proof: Let U be the non-empty class of unbiased estimators of θ. Let $h_1, h_2 \in U$. Since T is sufficient for θ, it follows by the Rao-Blackwell Theorem that $E(h_1|T), E(h_2|T) \in U$. Moreover,

$$\theta = E(E(h_1|T)) = E(E(h_2|T)) \qquad \Longleftrightarrow \qquad E(E(h_1|T) - E(h_2|T)) = 0.$$

Since T is complete, it follows that $E(h_1|T) = E(h_2|T)$. But, $h_1, h_2 \in U$ are arbitrary, so that $E(h|T)$ is the same for all $h \in U$. Thus, by the Rao-Blackwell Theorem, $E(h|T)$ improves all estimators of U. Therefore, since $E(h|T) \in U$, then $E(h|T)$ is the unique UMVUE for θ. $\blacksquare$

PRINCIPLE PROPERTIES OF THE INVERSE GAUSSIAN DISTRIBUTION:

Let $X \sim \text{IG}(\mu, \lambda)$. Then, the pdf for X is given by

$$f_X(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\} \cdot 1_{\{x\in(0,\infty);\lambda,\mu>0\}}. \tag{1.1}$$

Letting $\frac{1}{2}\alpha^2 = \mu$, Tweedie [1], claims that (1.1) is equivalent to

$$f_1(x; \alpha, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{-\lambda(\alpha \cdot x - (2\alpha)^{1/2} + 1/2x)\right\}. \tag{1.2}$$

By inspection, a necessary and sufficient condition for (1.2) and (1.1) to be equivalent is that $\alpha = (2\mu^2)^{-1}$. Thus, the ***substitution of Tweedie***, namely $\frac{1}{2}\alpha^2 = \mu$, is ***invalid*** for (1.2) and (1.1) to be equivalent.

Moment Generating Function of X : Assume that $\alpha = (2\mu^2)^{-1}$, so that (1.2) and (1.1) are equivalent. To obtain the moment generating function (MGF) of X, Tweedie introduces the notion of the Laplace transform, $L_X(t; \mu, \lambda) = \log(E(e^{-t\cdot X}))$. Note that $L_X(-t; \mu, \lambda)$, if it exists, is the cumulant generating function for the random variable X. Subsequently, if it exists, the MGF for the random variable X is given by $\exp\{L_X(-t; \mu, \lambda)\}$. Suppose that $t \in \mathbb{C}$, where $\mathbb{C}$ is the set of complex numbers. Tweedie shows that

$$L_X(t; \alpha, \lambda) = \lambda(2\alpha)^{1/2} - \lambda\sqrt{2}\left(\alpha + \frac{t}{\lambda}\right)^{1/2} + \log\left(\int_0^\infty f_1(x; \alpha + \frac{t}{\lambda}, \lambda)dx\right), \tag{1.3}$$

and provided that: (i) $\text{Re}(t) = 0$; or (ii) $\text{Re}(t) > -\alpha \cdot \lambda$, the integral of (1.3) evaluates to unity. Assume now that $\text{Im}(t) = 0$, so that $t \in \mathbb{R}$. Then, provided that $t > -\alpha \cdot \lambda$, since $\log(1) = 0$, (1.3) reduces to

$$\begin{aligned} L_X(t; \alpha, \lambda) &= \lambda(2\alpha)^{1/2} - \lambda\sqrt{2}\left(\alpha + \frac{t}{\lambda}\right)^{1/2} \\ &= \frac{\lambda}{\mu} - \frac{\lambda\sqrt{2}}{\sqrt{2}\mu}\left(1 + \frac{2\mu^2 t}{\lambda}\right)^{1/2} \\ &= \frac{\lambda}{\mu}\left(1 - \left(1 + \frac{2\mu^2 t}{\lambda}\right)^{1/2}\right). \end{aligned} \tag{1.4}$$

Thus, provided that $-t < \alpha \cdot \lambda$, the MGF for X, is

$$M_X(t) = \exp\{L_X(-t; \mu, \lambda)\} = \exp\left\{\frac{\lambda}{\mu}\left(1 - \left(1 - \frac{2\mu^2 t}{\lambda}\right)^{1/2}\right)\right\} \cdot 1_{\left\{t<\frac{\lambda}{2\mu^2}\right\}}. \square$$

Characteristic Function of X : The result immediately follows from (1.3) and the note following the expression. This is due to the fact that $\Phi_X(t) = \exp\{L_X(-i \cdot t; \mu, \lambda)\}$, and $\mathrm{Re}(i \cdot t) = 0$. Hence, from (1.3), and (1.4), it is

$$\Phi_X(t) = \exp\left\{L_X(-i \cdot t; \mu, \lambda)\right\} = \exp\left\{\frac{\lambda}{\mu}\left(1 - \left(1 - \frac{2\mu^2 i \cdot t}{\lambda}\right)^{1/2}\right)\right\}.\square$$

f **Defines a Pdf:** By inspection, it is clear that (1.1) is positive for all $x, \lambda, \mu \in (0, \infty)$. Shuster [2], shows that the cdf for X is given by

$$F_X(x) = \Phi[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)] + \exp\left\{\frac{2\lambda}{\mu}\right\}\Phi[-\sqrt{\frac{\lambda}{x}}\left(1 + \frac{x}{\mu}\right)], \tag{1.5}$$

where Φ is the cdf for the standard normal distribution. It holds,

$$\lim_{x \to 0} F_X(x) = \lim_{x \to -\infty} \Phi(x) + \exp\left\{\frac{2\lambda}{\mu}\right\} \lim_{x \to -\infty} \Phi(x) = 0; \text{ and}$$
$$\lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} \Phi(x) + \exp\left\{\frac{2\lambda}{\mu}\right\} \lim_{x \to -\infty} \Phi(x) = 1,$$

where the later limit can be shown using L'Hospital's Rule. Therefore, f, as defined by (1.1), defines a valid pdf.$\square$

The Pdf of the Reciprocal of X : Let $Y = \frac{1}{X}$. Since the support of X is $(0, \infty)$, the transformation is one-to-one. Also, note that Y is monotone (decreasing) for $X \in (0, \infty)$. It holds,

$$\begin{aligned}
f_Y(y) &= |\frac{d}{dy} g^{-1}(y)| f_X(g^{-1}(y)) \cdot 1_{\{y \in (0,\infty)\}} \\
&= |-\frac{1}{y^2}| f_X(\frac{1}{y}) \cdot 1_{\{y \in (0,\infty)\}} \\
&= \frac{1}{y^2}\left(\frac{y^3\lambda}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda \cdot y(y^{-1} - \mu)^2}{2\mu^2}\right\} \cdot 1_{\{y,\mu,\lambda \in (0,\infty)\}}. \\
&= \left(\frac{\lambda}{2\pi \cdot y}\right)^{1/2} \exp\left\{-\frac{\lambda(1 - \mu \cdot y)^2}{2\mu^2 y}\right\} \cdot 1_{\{y,\mu,\lambda \in (0,\infty)\}}.\square
\end{aligned}$$

Expected Value and Variance of X **:** Let $\kappa(t) = \log(M_X(t))$. It holds

$$E(X) = \frac{d}{dt}\kappa(t)|_{t=0}; \text{ and}$$
$$\text{Var}(X) = \frac{d^2}{dt^2}\kappa(t)|_{t=0}. \tag{1.6}$$

Proof: It is,

$$\frac{d}{dt}\kappa(t) = \frac{d}{dt}\log(M_X(t)) = \frac{1}{M_X(t)}\frac{d}{dt}M_X(t),$$

so that

$$\frac{d}{dt}\kappa(t)|_{t=0} = \frac{1}{M_X(0)}M'_X(0) = E(X).$$

Also, by the product rule for derivatives, it holds

$$\frac{d^2}{dt^2}\kappa(t) = \frac{d}{dt}\left(\frac{1}{M_X(t)}M'_X(t)\right) = -\frac{M'_X(t)}{(M_X(t))^2}\frac{d}{dt}M_X(t) + \frac{1}{M_X(t)}M''_X(t),$$

so that

$$\frac{d^2}{dt^2}\kappa(t)|_{t=0} = -\frac{(M'_X(0))^2}{(M_X(0))^2} + \frac{1}{M_X(0)}M''_X(0) = \text{Var}(X).\square$$

From (1.4), it is

$$\kappa(t) = L_X(-t;\mu,\lambda) = \frac{\lambda}{\mu}\left(1 - \left(1 - \frac{2\mu^2 t}{\lambda}\right)^{1/2}\right) \cdot 1_{\left\{t<\frac{\lambda}{2\mu^2}\right\}}.$$

Hence,

$$\frac{d}{dt}\kappa(t) = -\frac{\lambda}{2\mu}\left(1 - \frac{2\mu^2 t}{\lambda}\right)^{-1/2}\left(-\frac{2\mu^2}{\lambda}\right) = \mu\left(1 - \frac{2\mu^2 t}{\lambda}\right)^{-1/2}; \text{ and}$$
$$\frac{d^2}{dt^2}\kappa(t) = -\frac{\mu}{2}\left(1 - \frac{2\mu^2 t}{\lambda}\right)^{-3/2}\left(-\frac{2\mu^2}{\lambda}\right) = \frac{\mu^3}{\lambda}\left(1 - \frac{2\mu^2 t}{\lambda}\right)^{-3/2}.$$

Therefore, from (1.6), it holds that $E(X) = \mu$, and $\text{Var}(X) = \frac{\mu^3}{\lambda}$.$\square$

Sufficient and Complete Statistic: We re-write (1.2) as

$$\log(f_1(x;\alpha,\lambda)) = D(\lambda,\mu) + S(x) + \sum_{i=1}^{2} Q_i(\lambda,\mu)T_i(x), \tag{1.7}$$

where

$$D(\lambda,\mu) = \frac{1}{2}[\log(\frac{\lambda}{2\pi}) + 2\lambda\sqrt{2\alpha}]; \quad S(x) = -\frac{3}{2}\log(x);$$
$$Q_1(\lambda,\mu) = -\lambda\cdot\alpha; \quad T_1(x) = x;$$
$$Q_2(\lambda,\mu) = -\lambda; \text{ and } T_2(x) = \frac{1}{x}.$$

It follows from (1.7),

$$f_1(x;\alpha,\lambda) = \exp\left\{D(\lambda,\mu) + S(x) + \sum_{i=1}^{2} Q_i(\lambda,\mu)T_i(x)\right\},$$

so that f_1 belongs to a 2-dimensional parameter exponential family. Therefore, (via several Theorems), $(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} \frac{1}{X_i})$ are jointly sufficient/complete for (μ,λ).□

Maximum Likelihood Estimators of (μ,λ) **:** Let $X_1,\ldots,X_n$ be a random sample from IG(μ,λ), let $L(\mu,\lambda;\boldsymbol{x})$ be the likelihood function, and let LL$(\mu,\lambda;\boldsymbol{x})$ be the log of the likelihood function. It follows from (1.1), that

$$\begin{aligned} L(\mu,\lambda;\boldsymbol{x}) &= f\left(x_1,\ldots,x_n;\mu,\lambda\right) \\ &= \prod_{i=1}^{n}\left(\frac{\lambda}{2\pi\cdot x_i^3}\right)^{1/2}\exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}\cdot 1_{\{x_i\in(0,\infty);\lambda,\mu>0\}} \\ &= \left(\frac{\lambda}{2\pi}\right)^{n/2}\prod_{i=1}^{n} x_i^{-3/2}\exp\left\{-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}\right\}\cdot 1_{\{x_i\in(0,\infty);\lambda,\mu>0\}}, \end{aligned}$$

from which,

$$\mathrm{LL}(\mu,\lambda;\boldsymbol{x}) = \frac{n}{2}\left(\log(\lambda)-\log(2\pi)\right) - \frac{3}{2}\sum_{i=1}^{n}\log(x_i) - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i}.$$

Differentiating the log-likelihood function ($w.r.t.\ \mu$), it is

$$\begin{aligned} \frac{\partial}{\partial\mu}\mathrm{LL}(\mu,\lambda;\boldsymbol{x}) &= \frac{\lambda}{\mu^3}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i} + \frac{\lambda}{\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)}{x_i} \\ &= \frac{\lambda}{\mu^3}\left(n\cdot\overline{x} - 2n\cdot\mu + \mu^2\sum_{i=1}^{n}\frac{1}{x_i}\right) + \frac{n\cdot\lambda}{\mu^2} - \frac{\lambda}{\mu}\sum_{i=1}^{n}\frac{1}{x_i} \\ &= n\cdot\overline{x}\frac{\lambda}{\mu^3} - \frac{n\cdot\lambda}{\mu^2}. \end{aligned} \tag{1.8}$$

Setting (1.8) equal to zero, we obtain that $\hat{\mu} = \overline{X}$. Next, differentiating the log-likelihood function ($w.r.t\ \lambda$), it is

$$\begin{aligned} \frac{\partial}{\partial\lambda}\mathrm{LL}(\mu,\lambda;\boldsymbol{x}) &= \frac{n}{2\lambda} - \frac{1}{2\mu^2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{x_i} \\ &= \frac{n}{2\lambda} - \frac{1}{2\mu^2}\sum_{i=1}^{n}\left(x_i - 2\mu + \frac{\mu^2}{x_i}\right) \\ &= \frac{n}{2\lambda} - \frac{n\cdot\overline{x}}{2\mu^2} + \frac{n}{\mu} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{x_i}. \end{aligned} \tag{1.9}$$

Setting (1.9) equal to zero, and assigning μ to be its MLE, $\hat{\mu} = \overline{X}$, we have

$$\begin{aligned}
& 0 = \frac{n}{2\lambda} - \frac{n \cdot \overline{x}}{2(\overline{x})^2} + \frac{n}{\overline{x}} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{x_i} \\
\iff & -\frac{n}{2\lambda} = \frac{n}{2\overline{x}} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{x_i} \\
\iff & \frac{1}{\lambda} = -\frac{1}{\overline{x}} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} \\
\iff & \frac{1}{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i} - \frac{1}{\overline{x}}\right). \qquad (1.10)
\end{aligned}$$

Therefore, the maximum likelihood estimators for (μ, λ) are $\hat{\mu} = \overline{X}$ and $\hat{\lambda} = n \cdot \left(\sum_{i=1}^{n}(\frac{1}{x_i} - \frac{1}{\overline{x}})\right)^{-1}$. Of course, to verify that these estimates truly maximize the likelihood function, we would need to examine the second derivatives ("pure" and "partials" with respect to each of μ and λ), as well as the Jacobian (Wronskian) of the second partial derivatives. In particular, the signs of these derivatives.□

UMVUE for λ : Let $X_1, ..., X_n$ be a random sample from IG(μ, λ). In his paper, Tweedie shows that

$$\frac{1}{\hat{\lambda}} \sim \frac{\chi^2_{n-1}}{n \cdot \lambda}.$$

Recall, the expected value for a chi-square random variable is its respective degrees-of-freedom. Thus,

$$E\left(\frac{1}{\hat{\lambda}}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{x_i} - \frac{1}{\overline{x}}\right)\right) = E\left(\frac{Y}{n \cdot \lambda}\right) = \frac{n-1}{n \cdot \lambda},$$

where $Y \sim \chi^2_{n-1}$. It follows that $\frac{n}{n-1} \cdot \frac{1}{\hat{\lambda}}$ is unbiased for $\frac{1}{\lambda}$. Thus, provided it is unbiased for λ, then $\frac{n-1}{n}\hat{\lambda}$ is UMVUE for λ by the Lehmann-Scheffée Theorem, since this estimator is a function of the joint sufficient/complete statistics, $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n}\frac{1}{X_i}\right)$. Hence, it suffices to show that

$$E\left(\frac{n-1}{n}\hat{\lambda}\right) = \lambda. \qquad (1.11)$$

To show this, we require the result of the following Lemma:

Lemma: Let $X \sim \Gamma(n, \lambda)$, and let $Y = X^{-1}$. Then, $E(Y) = \lambda \cdot n^{-1}$.

Proof: Let $Y = g(X) = X^{-1}$. Then, $X = g^{-1}(Y) = Y^{-1}$. The transformation is one-to-one, since the support of X is $(0, \infty)$, and Y is monotone (decreasing) on this interval. Hence,

$$\begin{aligned} f_Y(y) &= |\frac{d}{dy} g^{-1}(y)| f_X(g^{-1}(y)) \cdot 1_{\{y \in (0,\infty)\}} \\ &= |-\frac{1}{y^2}| f_X(\frac{1}{y}) \\ &= \frac{\lambda^n}{y^{n+1}\Gamma(n)} \exp\left\{-\lambda \cdot y^{-1}\right\} \cdot 1_{\{y \in (0,\infty)\}}. \end{aligned}$$

Now, substituting $u = \lambda \cdot y^{-1}$, it is

$$\begin{aligned} E(Y) &= \int_0^\infty \frac{y \cdot \lambda^n}{y^{n+1}\Gamma(n)} \exp\left\{-\lambda \cdot y^{-1}\right\} dy \\ &= \frac{\lambda}{\Gamma(n)} \int_0^\infty u^{n-2} \exp\{-u\} \, dy \\ &= \frac{\Gamma(n-1)\lambda}{\Gamma(n)} \\ &= \lambda \cdot n^{-1}, \end{aligned}$$

as desired. Therefore, if $X \sim \Gamma(n, \lambda)$, then $E(Y) = E(\frac{1}{X}) = \lambda \cdot n^{-1}$.□
Now, let $Y \sim \chi^2_{n-1}$. Then, $Y \sim \Gamma(\frac{n-1}{2}, \frac{1}{2})$, so that $E(\frac{1}{Y}) = \frac{1}{n-1}$. Hence, from (1.10), it holds

$$E\left(\frac{n-1}{n}\hat{\lambda}\right) = E\left((n-1)\,\frac{n \cdot \lambda}{n \cdot Y}\right) = (n-1)\,\lambda \cdot E\left(\frac{1}{Y}\right) = \lambda.$$

Therefore, by the Lehmann-Sheffée Theorem, $(n-1) \cdot \left(\sum_{i=1}^{n} (\frac{1}{x_i} - \frac{1}{\bar{x}})\right)^{-1}$ is the UMVUE for λ.□

Cramer-Rao Lower Bound for λ $(\mu = 1)$ **:** Let $X_1, ..., X_n$ be a random sample from IG(μ, λ). It is desired to determine the CRLB for λ when $\mu = 1$. To determine the CRLB, we consider the derivative of the log of the density function, as given by (1.1) above. Assume that the regularity conditions of the theorem hold. It is,

$$\log(f(x; \mu, \lambda)) = \frac{1}{2}[\log(\lambda) - \log(2\pi \cdot x^3)] - \frac{\lambda(x-1)^2}{2x},$$

which implies that

$$\frac{\partial}{\partial \lambda} \log(f(x; \mu, \lambda)) = \frac{1}{2\lambda} - \frac{(x-1)^2}{2x}. \tag{1.12}$$

Next, we need to determine the expectation ($w.r.t\ \lambda$) of the square of (1.12). However, since f is an exponential family, Lemma 7.3.11 of Casella states that

$$E_\lambda\left(\left(\frac{\partial}{\partial\lambda}\log(f(x;\mu,\lambda))\right)^2\right) = -E_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log(f(x;\mu,\lambda))\right).$$

Thus, we have

$$\frac{\partial^2}{\partial\lambda^2}\log(f(x;\mu,\lambda)) = -\frac{1}{2\lambda} \qquad \Longleftrightarrow \qquad -E_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log(f(x;\mu,\lambda))\right) = (2\lambda)^{-1}.$$

Since $\psi(\lambda) = \lambda$, then $\psi'(\lambda) = 1$. Hence, the CRLB for the variance of any unbiased estimator of λ, $T(\boldsymbol{X})$, is

$$\begin{aligned}\operatorname{Var}_\lambda(T(\boldsymbol{X})) &\geq \frac{(\psi'(\lambda))^2}{n \cdot E_\lambda\left(\left(\frac{\partial}{\partial\lambda}\log\left(f\left(x;\mu,\lambda\right)\right)\right)^2\right)} \\ &= \frac{1}{-n \cdot E_\lambda(\frac{\partial^2}{\partial\lambda^2}\log(f(x;\mu,\lambda)))} \\ &= \frac{2\lambda}{n}.\end{aligned}$$

Therefore, the CRLB (provided that the regularity conditions hold) for $\operatorname{Var}_\lambda(T(\boldsymbol{X}))$ is $2\lambda \cdot n^{-1}$.□

Convolutions of IG Random Variables: Let $X_1, ..., X_n$ be a random sample from $\mathrm{IG}(\mu,\lambda)$. It holds that $\overline{X}$ and $Y = a \cdot X$ ($a > 0$) also follow IG distributions. Let's determine the parametrization for these random variables. It follows that

$$\begin{aligned}E(\overline{X}) &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu; \text{ and} \\ \operatorname{Var}(\overline{X}) &= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = \frac{\mu^3}{n\cdot\lambda}.\end{aligned}$$

Thus, $\overline{X} \sim \mathrm{IG}(\mu, n\cdot\lambda)$. Next, we consider the random variable, $Y = a \cdot X$. Here,

$$\begin{aligned}E(Y) &= a \cdot E(X) = a\cdot\mu; \text{ and} \\ \operatorname{Var}(Y) &= a^2\operatorname{Var}(X) = \frac{a^2\mu^3}{\lambda} = \frac{(a\cdot\mu)^3}{a\cdot\lambda}.\end{aligned}$$

Thus, $Y \sim \mathrm{IG}(a\cdot\mu, a\cdot\lambda)$. Therefore, $\overline{X} \sim \mathrm{IG}(\mu, n\cdot\lambda)$ and $Y \sim \mathrm{IG}(a\cdot\mu, a\cdot\lambda)$. ■