Norwegian University of Science and Technology
Department of Mathematical Sciences

English

Contact during exam:
Arvid Naess 73 59 70 53/ 99 53 83 50

SOLUTIONS EXAM IN COURSE TMA4295 STATISTICAL INFERENCE

May 26, 2009
Time: 09:00–13:00

Permitted aids: *Tabeller og formler i statistikk*, Tapir Forlag
K. Rottmann: *Matematisk formelsamling*
Calculator HP30S / CITIZEN SR-270X
Yellow, stamped A5-sheet with your own handwritten notes.

Examination results are due: June 16, 2009

Problem 1

A statistical distribution often used to model income is the Pareto distribution. The associated probability density function (pdf) is given as,

$$f(x|\theta) = \begin{cases} \frac{1}{\theta} c^{\frac{1}{\theta}} \, x^{-(1+\frac{1}{\theta})} & \text{for } x > c\,, \\ \\ 0 & \text{for } x \le c\,, \end{cases}$$

where θ is an unknown parameter and $c > 0$ is a known constant.

If the random variable X has the pdf $f(x|\theta)$ given above, then it can be shown that,

$$\mathrm{E}_\theta[X^k] = \begin{cases} \frac{c^k}{1-k\theta} & \text{for } k < \frac{1}{\theta}\,, \\ \\ \infty & \text{for } k \ge \frac{1}{\theta}\,, \end{cases}$$

Let $X_1, \ldots, X_n$ be a random sample from the given Pareto distribution.

a) Determine the method of moments estimator (MME) Θ^* of θ based on the random sample $X_1, \ldots, X_n$.

Prove that Θ^* is consistent for a suitable condition on θ.

SOLUTION:

Since $\mu_1 = \mathrm{E}[X] = \frac{c}{1-\theta}$ (for $\theta < 1$), and $m_1 = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}$ is the empirical mean from a realization $x_1, \ldots, x_n$ of the random sample, the equation to find the MME will be

$$m_1 = \frac{c}{1-\theta^*}\,, \text{that is}\,, \theta^* = 1 - \frac{c}{\overline{x}}$$

This gives the MME

$$\Theta^* = 1 - \frac{c}{\overline{X}}$$

By the WLLN (weak law of large numbers), if $\mu_2 = \mathrm{E}[X^2] < \infty$, that is $\theta < 1/2$, then $\overline{X} \xrightarrow[n\to\infty]{P} \mu_1 = \frac{c}{1-\theta}$. The function $g(x) = 1 - \frac{c}{x}$ is continuous on $(0, \infty)$. Then $\overline{X} \xrightarrow[n\to\infty]{P} \mu_1$ implies that $\Theta^* = g(\overline{X}) \xrightarrow[n\to\infty]{P} g(\mu_1) = 1 - \frac{c}{\mu_1} = 1 - \frac{c}{c/(1-\theta)} = \theta$. Hence, Θ^* is consistent if $\theta < 1/2$. (Actually, the WLLN is true here if $\mathrm{E}[X] < \infty$, that is, if $\theta < 1$.)

b) Establish the approximate distribution for Θ^* for large values of n.

For which values of θ is the approximation valid?

SOLUTION:

By the CLT (central limit theorem), $\sqrt{n}\big(\overline{X} - \mu_1\big) \xrightarrow[n\to\infty]{d} N(0, \sigma^2)$ (convergence in distribution) if $\mathrm{Var}[X_i] = \sigma^2 < \infty$, that is $\theta < 1/2$. Invoking the (first-order) Delta Method for the function $g(x) = 1 - \frac{c}{x}$, if $g'(\mu_1)$ exists and is non-zero, it is obtained that

$$\sqrt{n}\big(g(\overline{X}) - g(\mu_1)\big) \xrightarrow[n\to\infty]{d} N(0, \sigma^2\, g'(\mu_1)^2)\,,$$

that is,

$$\sqrt{n}\big(\Theta^* - \theta\big) \xrightarrow[n\to\infty]{d} N(0, \sigma^2\, g'(\mu_1)^2)\,,$$

The function g is differentiable on $(0, \infty)$, and $g'(x) = \frac{c}{x^2}$. This gives $g'(\mu_1) = \frac{c}{\mu_1^2} = \frac{c}{c^2/(1-\theta)^2} = \frac{(1-\theta)^2}{c} > 0$ for $\theta < 1$.

$$\sigma^2 = \mu_2 - \mu_1^2 = \frac{c^2}{1-2\theta} - \frac{c^2}{(1-\theta)^2} = \frac{c^2\theta^2}{(1-2\theta)(1-\theta)^2}\,,$$

which leads to

$$\sigma^2\, g'(\mu_1)^2 = \frac{\theta^2(1-\theta)^2}{(1-2\theta)}\,.$$

Hence, for large values of n,

$$\Theta^* \approx Y_n \sim N\Big(\theta, \frac{\theta^2(1-\theta)^2}{n(1-2\theta)}\Big),$$

which is guaranteed if $\theta < 1/2$.

c) Assume that $X \sim f(x|\theta)$, and define

$$U = \ln X - \ln c\,.$$

Show that U is exponentially distributed with pdf

$$f_U(u|\theta) = \begin{cases} \frac{1}{\theta}\, e^{-\frac{u}{\theta}} & \text{for } u > 0\,, \\ 0 & \text{for } u \leq 0\,, \end{cases}$$

Use this result, or possibly some other way, to derive Cramér-Rao's lower bound on the variance of unbiased estimators of θ. You do not need to verify the conditions for the validity of the Cramér-Rao theorem (interchange of expectation and derivation).

SOLUTION:

For $u > 0$, that is $x > c$,

$$f_U(u|\theta) = f(ce^u|\theta)|dce^u/du| = f(ce^u|\theta)ce^u = \frac{1}{\theta}c^{\frac{1}{\theta}}\,(ce^u)^{-(1+\frac{1}{\theta})}ce^u = \frac{1}{\theta}\,e^{-\frac{u}{\theta}}\,,\ (u>0)\,.$$

For $u \leq 0$, that is, $x \leq c$, clearly, $f_U(u|\theta) = 0$.

Due to the one-to-one character of the mapping $X \to U$, any unbiased estimator $T(\mathbf{X})$ ($\mathbf{X} = (X_1, \ldots, X_n)$) is equivalent with an unbiased estimator $W(\mathbf{U})$. Hence,

$$\mathrm{Var}[T(\mathbf{X})] = \mathrm{Var}[W(\mathbf{U})] \geq \frac{1}{nI_0(\theta)}\,,$$

where the Fisher information $I_0(\theta)$ is given as,

$$I_0(\theta) = \mathrm{E}_\theta\Big[\Big(\frac{\partial}{\partial\theta}\ln f_U(U|\theta)\Big)^2\Big]\,,$$

Now,

$$\Big(\frac{\partial}{\partial\theta}\ln f_U(U|\theta)\Big)^2 = \Big(-\frac{1}{\theta} + \frac{U}{\theta^2}\Big)^2 = \frac{1}{\theta^2} - \frac{2}{\theta^3}U + \frac{1}{\theta^4}U^2\,,$$

which gives,

$$I_0(\theta) = \mathrm{E}_\theta\Big[\Big(\frac{\partial}{\partial\theta}\ln f_U(U|\theta)\Big)^2\Big] = \frac{1}{\theta^2}\Big(1 - \frac{2}{\theta}\theta + \frac{1}{\theta^2}2\theta^2\Big) = \frac{1}{\theta^2}\,.$$

Hence,

$$\mathrm{Var}[T(\mathbf{X})] = \mathrm{Var}[W(\mathbf{U})] \geq \frac{\theta^2}{n}\,.$$

d) Show that the maximum likelihood estimator $\hat{\Theta}$ of θ based on the random sample $X_1, \ldots, X_n$ is given by

$$\hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n} \ln X_i - \ln c\,.$$

Is $\hat{\Theta}$ an UMVUE (uniform minimum variance unbiased estimator)?

SOLUTION:

We may calculate the MLE using the random sample $U_1, \ldots, U_n$. For a realization $u_1, \ldots, u_n$, the loglikelihood function $l(\theta|u_1, \ldots, u_n)$ becomes,

$$l(\theta|u_1, \ldots, u_n) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} u_i\,.$$

$$dl(\theta|u_1, \ldots, u_n)/d\theta = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} u_i = 0 \implies \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} u_i.$$

Hence, the MLE $\hat{\Theta}$ of θ is given as,

$$\hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n} U_i = \frac{1}{n}\sum_{i=1}^{n} \ln X_i - \ln c.$$

It is found that

$$\mathrm{E}[\hat{\Theta}] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[U_i] = \frac{1}{n}\sum_{i=1}^{n} \theta = \theta\,,$$

so that $\hat{\Theta}$ is an unbiased estimator of θ. Also,

$$\mathrm{Var}[\hat{\Theta}] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[U_i] = \frac{1}{n^2}\sum_{i=1}^{n} \theta^2 = \frac{\theta^2}{n}\,.$$

Comparing with the result from c), it follows that $\hat{\Theta}$ is an efficient estimator of θ. Hence, we may conclude that $\hat{\Theta}$ is UMVUE.

e) Derive an exact $1-\alpha$ confidence interval for θ based on the observations $x_1, \ldots, x_n$ of the random sample $X_1, \ldots, X_n$. (Hint: Find the pdf of $2U/\theta$.) If you are unable derive the exact confidence interval, you will get half credit for an approximate $1-\alpha$ confidence interval.

SOLUTION:

The RV $V = 2U/\theta$ has the pdf $f_V(v) = \frac{1}{2}\, e^{-\frac{v}{2}}$ $(v > 0)$, which shows that V a χ^2 variable with 2 degrees of freedom. By the addition theorem for independent χ^2 variables,

$$\frac{2n\hat{\Theta}}{\theta} = \sum_{i=1}^{n} V_i$$

becomes a χ^2 variable with $2n$ degrees of freedom χ^2_{2n}. Therefore $(\text{Prob}(\chi^2_{2n} \leq \chi^2_{2n,\gamma}) = 1 - \gamma)$,

$$\text{Prob}\Big(\chi^2_{2n,1-\frac{\alpha}{2}} \leq \frac{2n\hat{\Theta}}{\theta} \leq \chi^2_{2n,\frac{\alpha}{2}}\Big) = 1 - \alpha\,,$$

which can be written as,

$$\text{Prob}\Big(\frac{2n\hat{\Theta}}{\chi^2_{2n,\frac{\alpha}{2}}} \leq \theta \leq \frac{2n\hat{\Theta}}{\chi^2_{2n,1-\frac{\alpha}{2}}}\Big) = 1 - \alpha\,.$$

This leads to the following exact $1 - \alpha$ confidence interval for θ based on a realization $x_1, \ldots, x_n$ of the random sample $X_1, \ldots, X_n$:

$$\Big[\frac{2\sum_{i=1}^{n} \ln x_i - 2n \ln c}{\chi^2_{2n,\frac{\alpha}{2}}}, \frac{2\sum_{i=1}^{n} \ln x_i - 2n \ln c}{\chi^2_{2n,1-\frac{\alpha}{2}}}\Big]\,.$$

An approximate confidence interval can be derived from the asymptotic result that (CLT),

$$\frac{\sqrt{n}\big(\hat{\Theta} - \theta\big)}{\theta} \xrightarrow[n\to\infty]{d} N(0,1)\,.$$

Hence, for large n,

$$\text{Prob}\Big(-z_{\frac{\alpha}{2}} \leq \frac{\sqrt{n}\big(\hat{\Theta} - \theta\big)}{\theta} \leq z_{\frac{\alpha}{2}}\Big) \approx 1 - \alpha\,,$$

that is,

$$\text{Prob}\Big(\frac{\hat{\Theta}}{1 - z_{\frac{\alpha}{2}}/\sqrt{n}} \leq \theta \leq \frac{\hat{\Theta}}{1 + z_{\frac{\alpha}{2}}/\sqrt{n}}\Big) \approx 1 - \alpha\,,$$

This leads to the following approximate $1 - \alpha$ confidence interval for θ based on a realization $x_1, \ldots, x_n$ of the random sample $X_1, \ldots, X_n$ for large n:

$$\Big[\frac{\frac{1}{n}\sum_{i=1}^{n} \ln x_i - \ln c}{1 - z_{\frac{\alpha}{2}}/\sqrt{n}}, \frac{\frac{1}{n}\sum_{i=1}^{n} \ln x_i - \ln c}{1 + z_{\frac{\alpha}{2}}/\sqrt{n}}\Big]\,.$$

Problem 2

a) Let $X_1, \ldots, X_n$ be a random sample from a pdf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family given by

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, $d \leq k$.

By using the Factorization Theorem, prove that

$$T(\mathbf{X}) = \Big(\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\Big)$$

is a sufficient statistic for $\boldsymbol{\theta}$.

<u>SOLUTION:</u>

The sample joint pdf is given as

$$f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{j=1}^{n} f(x_j|\boldsymbol{\theta}) = \prod_{j=1}^{n} h(x_j)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x_j)\right),$$

$$= \Big(\prod_{j=1}^{n} h(x_j)\Big)(c(\boldsymbol{\theta}))^n \exp\left(w_1(\boldsymbol{\theta})\sum_{j=1}^{n} t_1(x_j) + \ldots + w_k(\boldsymbol{\theta})\sum_{j=1}^{n} t_k(x_j)\right),$$

$$= (c(\boldsymbol{\theta}))^n \exp\left(w_1(\boldsymbol{\theta})T_1(\mathbf{x}) + \ldots + w_k(\boldsymbol{\theta})T_k(\mathbf{x})\right) \Big(\prod_{j=1}^{n} h(x_j)\Big),$$

$$= (c(\boldsymbol{\theta}))^n \exp\big(w(\boldsymbol{\theta})T(\mathbf{x})^\top\big)\, \tilde{h}(\mathbf{x}),$$

where $T_i(\mathbf{x}) = \sum_{j=1}^{n} t_i(x_j)$ for $i = 1, \ldots, k$, $w(\boldsymbol{\theta}) = (w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta}))$, $T(\mathbf{x}) = (T_1(\mathbf{x}), \ldots, T_k(\mathbf{x}))$ and $\tilde{h}(\mathbf{x}) = \prod_{j=1}^{n} h(x_j)$.

Hence, it is clear that $f(\mathbf{x}|\boldsymbol{\theta}) = \tilde{g}(T(\mathbf{x}), \boldsymbol{\theta})\, \tilde{h}(\mathbf{x})$ for suitable functions $\tilde{g}$ and $\tilde{h}$. By the factorization theorem it now follows that $T(\mathbf{X}) = (T_1(\mathbf{X}), \ldots, T_k(\mathbf{X}))$ is then a sufficient statistic for $\boldsymbol{\theta}$.

b) Show that the inverse Gaussian pdf

$$f(x|\mu, \lambda) = \begin{cases} \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) & \text{for } x > 0, \\ 0 & \text{for } x \leq 0, \end{cases}$$

where μ and λ are positive constants, belongs to an exponential family. Use the result from point a) to find a sufficient statistic for $\boldsymbol{\theta} = (\mu, \lambda)$.

SOLUTION:

For $x > 0$,

$$f(x|\mu,\lambda) = x^{-3/2}\left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left(-\frac{\lambda}{2\mu^2}x + \frac{\lambda}{\mu} - \frac{\lambda}{2x}\right)$$
$$= x^{-3/2}\left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left(\frac{\lambda}{\mu}\right) \exp\left(-\frac{\lambda}{2\mu^2}x - \frac{\lambda}{2x}\right),$$

which is seen to belong to an exponential family with $k = 2$, $\boldsymbol{\theta} = (\mu, \lambda)$, $h(x) = x^{-3/2}$, $c(\boldsymbol{\theta}) = \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left(\frac{\lambda}{\mu}\right)$, $w_1(\boldsymbol{\theta}) = -\frac{\lambda}{2\mu^2}$, $w_2(\boldsymbol{\theta}) = -\frac{\lambda}{2}$, $t_1(x) = x$ and $t_2(x) = 1/x$.

According to point a), a sufficient statistic for $\boldsymbol{\theta} = (\mu, \lambda)$ is then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} X_j, \sum_{j=1}^{n} 1/X_j\right)$$

c) Show that the maximum likelihood estimators of the parameters μ and λ in b) are

$$\hat{M} = \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j\,,$$

and

$$\hat{\Lambda} = \frac{n}{\sum_{j=1}^{n}\left(\frac{1}{X_j} - \frac{1}{\overline{X}}\right)}\,.$$

SOLUTION:

For a realization $x_1, \ldots, x_n$ of the random sample, the likelihood function becomes,

$$L(\mu,\lambda|x_1,\ldots,x_n) = \prod_{i=1}^{n} x_i^{-3/2}(2\pi)^{-n/2}\lambda^{n/2} \exp\left(\frac{n\lambda}{\mu}\right) \exp\left(-\frac{\lambda}{2\mu^2}\sum_{i=1}^{n} x_i - \frac{\lambda}{2}\sum_{i=1}^{n} 1/x_i\right),$$

A (reduced) log-likelihood function is then,

$$l(\mu,\lambda|x_1,\ldots,x_n) = \frac{n}{2}\ln\lambda + \frac{n\lambda}{\mu} - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n} x_i - \frac{\lambda}{2}\sum_{i=1}^{n} 1/x_i\,.$$

$$\frac{dl(\mu,\lambda|x_1,\ldots,x_n)}{d\mu} = -\frac{n\lambda}{\mu^2} + \frac{\lambda}{\mu^3}\sum_{i=1}^{n} x_i = 0\,,$$

gives the point estimate

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}\,.$$

$$\frac{dl(\mu,\lambda|x_1,\ldots,x_n)}{d\lambda} = \frac{n}{2\lambda} + \frac{n}{\mu} - \frac{1}{2\mu^2}\sum_{i=1}^{n} x_i - \frac{1}{2}\sum_{i=1}^{n} 1/x_i = 0\,,$$

Substituting the point estimate $\hat{\mu} = \overline{x}$ for μ, provides the following equation to determine the point estimate $\hat{\lambda}$,

$$\frac{n}{2\hat{\lambda}} + \frac{n}{2\overline{x}} - \frac{1}{2}\sum_{i=1}^{n} 1/x_i = 0\,,$$

which leads to the result,

$$\hat{\lambda} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} - \frac{1}{\overline{x}}} = \frac{n}{\sum_{j=1}^{n}\left(\frac{1}{x_j} - \frac{1}{\overline{x}}\right)}\,.$$

Hence, it follows that the MLE will be given as,

$$\hat{M} = \overline{X}\,.$$

and

$$\hat{\Lambda} = \frac{n}{\sum_{j=1}^{n}\left(\frac{1}{X_j} - \frac{1}{\overline{X}}\right)}\,.$$

d) Prove that a one-to-one function of a sufficient statistic is a sufficient statistic. Use this property together with the result from b) to show that $T^*(\mathbf{X}) = \big(\hat{M}, \hat{\Lambda}\big)$ is also a sufficient statistic for (μ,λ).

SOLUTION:

Assume that $T(\mathbf{X})$ is a sufficient statistic, and define $T^*(\mathbf{x}) = r\big(T(\mathbf{x})\big)$ for all $\mathbf{x}$, where r is a one-to-one function with inverse r^{-1}. By the factorization theorem there exist functions g and h such that

$$f(\mathbf{x}|\boldsymbol{\theta}) = g(T(\mathbf{x}),\boldsymbol{\theta})\,h(\mathbf{x}) = g(r^{-1}\big(T^*(\mathbf{x})\big),\boldsymbol{\theta})\,h(\mathbf{x}) = g^*(T^*(\mathbf{x}),\boldsymbol{\theta})\,h(\mathbf{x})\,,$$

where $g^*(t,\boldsymbol{\theta}) = g\big(r^{-1}(t),\boldsymbol{\theta}\big)$. Again, by the factorization theorem, it follows that $T^*(\mathbf{X})$ is a sufficient statistic for $\boldsymbol{\theta}$.

$T(\mathbf{X}) = \Big(\sum_{j=1}^{n} X_j, \sum_{j=1}^{n} 1/X_j\Big)$ is sufficient by b). Obviously $T_1(\mathbf{X}) = \Big(\frac{1}{n}\sum_{j=1}^{n} X_j, \frac{1}{n}\sum_{j=1}^{n} 1/X_j\Big) = \Big(\overline{X}, \frac{1}{n}\sum_{j=1}^{n} 1/X_j\Big)$ is obtained from $T(\mathbf{X})$ by a one-to-one mapping. Then, define $T_2(\mathbf{X}) = \Big(\overline{X}, \frac{1}{n}\sum_{j=1}^{n} 1/X_j - \frac{1}{\overline{X}}\Big) = r_1\big(T_1(\mathbf{X})\big)$. Assume that $r_1(x,y) =$

$r_1(u,v)$, which gives $x = u$ and $y - 1/x = v - 1/u = v - 1/x$, and therefore $y = v$. Hence, r_1 is a one-to-one mapping. Finally, define $T_3(\mathbf{X}) = r_2\big(T_2(\mathbf{X})\big)$, where $r_2(x,y) = (x, 1/y)$. Clearly, r_2 is a one-to-one mapping. It is seen that $T_3(\mathbf{X}) = \big(\hat{M}, \hat{\Lambda}\big)$. Since $T_3(\mathbf{X})$ is obtained by a sequence of one-to-one mappings of $T(\mathbf{X})$, which is then also one-to-one, $T_3(\mathbf{X})$ is therefore also sufficient.

e) Use Jensen's inequality, or any other suitable method, to verify that the estimator $\hat{\Lambda}$ in c) is positive and finite (a.s.). That is, verify that $\frac{1}{n}\sum_{j=1}^{n}\big(\frac{1}{X_j} - \frac{1}{\bar{X}}\big) > 0$, or, $\frac{1}{n}\sum_{j=1}^{n}\frac{1}{X_j} > \frac{1}{\bar{X}}$.

Jensen's inequality: If $g(\cdot)$ is a strictly convex function on the value range of a non-constant random variable Y, $0 < \mathrm{E}[|Y|] < \infty$, then $\mathrm{E}[g(Y)] > g(\mathrm{E}[Y])$. (Note: $g(x) = 1/x$ is a strictly convex function on $(0,\infty)$.)

SOLUTION:

We need to prove that $\frac{1}{n}\sum_{j=1}^{n}\frac{1}{x_j} > \frac{1}{\bar{x}}$ for any realization $x_1,\ldots,x_n$ of the random sample. Define the discrete random variable Y with values $x_1,\ldots,x_n$ by $\mathrm{Prob}(Y = x_j) = 1/n$, $j = 1,\ldots,n$. $Y > 0$ and non-constant (a.s.). Since $g(x) = 1/x$ is a strictly convex function on $(0,\infty)$, it follows from the Jensen inequality above that

$$\mathrm{E}\Big[\frac{1}{Y}\Big] = \sum_{j=1}^{n}\frac{1}{x_j}\frac{1}{n} > \frac{1}{\mathrm{E}[Y]} = \frac{1}{\sum_{j=1}^{n} x_j \frac{1}{n}} \quad \text{(a.s.)}\,,$$

which is exactly what we needed to prove.

An alternative proof by Cauchy-Schwartz inequality:

$$n^2 = \Big(\sum_{j=1}^{n}\frac{1}{\sqrt{x_j}}\sqrt{x_j}\Big)^2 \le \Big(\sum_{j=1}^{n}\frac{1}{x_j}\Big)\Big(\sum_{j=1}^{n} x_j\Big)\,,$$

which is seen to give the desired result if the inequality is strict. Equality can only occur if $x_i = c/x_i$ for every $i = 1,\ldots,n$, for some constant $c > 0$, that is, $x_i = \sqrt{c}$ for $i = 1,\ldots,n$. But this is impossible (a.s.), so the inequality is strict.