Statistica Neerlandica (1996) Vol. 50, nr. 2, pp. 325–332

Problem Section

Solutions

Problem 316 (V. Mammitzsch, Marburg, Germany)
Is the distribution function of $f(X+Y)-X$ continuous, if X and Y are independent and have continuous distribution functions?

This problem is of the type that it can be shortly formulated. However, the answer appeared to be difficult to the readers. No solution was submitted. The answer is, unfortunately, negative. A counterexample is not simply given. We refer to

Mammitzsch, V. (1980), Is the distribution function of $f(X+Y)-X$ continuous, if X and Y are independent and have continuous distribution functions?; Colloquia Mathematica Societatis János Bolya, vol. 32, Nonparametric Statistical Inference, Budapest.

Problem 317 (N. van Giersbergen, J. Goeree)
Let $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ be two random samples from an $\mathcal{N}(0,1)$ distribution. $P(n)$ denotes the probability that the second moment of the sample $\{X_i \pm Y_i\}_{i=1}^n$ is smaller than or equal to the second moment of the sample $\{Y_i\}_{i=1}^n$:

$$P(n) \equiv P\left[\sum_{i=1}^{n}(X_i \pm Y_i)^2 \leq \sum_{i=1}^{n} Y_i^2\right]$$

Prove for $n \geq 1$ that

$$P(2n-1) = \frac{1}{2} - \frac{1}{\pi} \arcsin\left(\frac{1}{\sqrt{5}}\right) - \frac{1}{4\pi}\sum_{k=1}^{n-1}\frac{1}{k}\binom{2k}{k}^{-1}\left(\frac{16}{5}\right)^k$$

$$P(2n) = \frac{1}{2} - \frac{1}{2\sqrt{5}} - \frac{1}{2}\sum_{k=1}^{n-1}\binom{2k}{k}\left(\frac{1}{5}\right)^{k+\frac{1}{2}}$$

No solutions were submitted, so we publish the solution by the proposers.

Solution by the proposers
Let $X = (X_1, \ldots, X_n)'$ and $Y = (Y_1, \ldots, Y_n)'$.
Define

$$J_1 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

$J_n = J_1 \otimes I_n$ and $Z' = (X', Y')$. Let D_1 be the orthogonal matrix which diagonalizes J_1, i.e.

$$D_1' J_1 D_1 = \Lambda_1,$$

where $\Lambda_1 = \mathrm{diag}(\lambda_1, \lambda_2)$ with $\lambda_1 = ½(1+\sqrt{5})$ and $\lambda_2 = ½(1-\sqrt{5})$. Writing $D_n = D_1 \otimes I_n$ and $W = D_n' Z$, we have $D_n' J_n D_n = \Lambda_1 \otimes I_n$ and $W \sim \mathcal{N}(0, I_{2n})$. One obtains

$$\begin{aligned} P[\| X+Y \| \le \| Y \|] &= P[X'X + 2X'Y \le 0] \\ &= P[Z' J_n Z \le 0] \\ &= P[W'(\Lambda_1 \otimes I_n) W \le 0] \end{aligned}$$

which may be further reduced to

$$\begin{aligned} P[\| X+Y \| \le \| Y \|] &= P\left[\frac{\sum_{i=1}^{n} W_i^2}{\sum_{i=n+1}^{2n} W_i^2} \le -\lambda_2/\lambda_1 \right] \\ &= P\left[F_{n,n} \le \frac{\sqrt{5}-1}{\sqrt{5}+1} \right] \\ &= I_{\frac{1}{2}-\frac{\sqrt{5}}{10}}\left(\frac{n}{2}, \frac{n}{2}\right) = \frac{1}{2} I_{4/5}\left(\frac{n}{2}, \frac{1}{2}\right) \end{aligned} \tag{1}$$

where $F_{n,n}$ is a random variable having an F-distribution with (n,n) degrees of freedom and $I_x(a,b)$ is the incomplete Beta-function defined as

$$I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t$$

In the last line of (1) we used: i) $P[F_{a,b} \le x] = I_{1-y}(a/2, b/2)$ with $y = b/(b+ax)$ and ii) $I_x(a,a) = ½ I_{1-y}(a, ½)$ with $y = 4(x-½)^2$. Using

$$I_x(a,b) = I_x(a-1,b) - \frac{\Gamma(a+b-1)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^b$$

and

$$\Gamma\left(\frac{1}{2}(2k+1)\right) = \sqrt{\pi} \binom{2k}{k} k! \left(\frac{1}{4}\right)^k$$

we find

$$\begin{aligned} P(2n-1) &= \frac{1}{2} I_{4/5}\left(\frac{2n-1}{2}, \frac{1}{2}\right) \\ &= \frac{1}{2} I_{4/5}\left(\frac{1}{2}, \frac{1}{2}\right) - \frac{1}{2} \sum_{k=1}^{n-1} \frac{\Gamma(k)}{\Gamma(\frac{2k+1}{2})\Gamma(\frac{1}{2})} \left(\frac{4}{5}\right)^{k-\frac{1}{2}} \left(\frac{1}{5}\right)^{\frac{1}{2}} \\ &= \frac{1}{2} - \frac{1}{\pi} \arcsin\left(\frac{1}{\sqrt{5}}\right) - \frac{1}{4\pi} \sum_{k=1}^{n-1} \frac{1}{k} \binom{2k}{k}^{-1} \left(\frac{16}{5}\right)^k \end{aligned}$$

and similarly

$$\begin{aligned}P(2n) &= \frac{1}{2}I_{4/5}\left(n,\frac{1}{2}\right)\\ &= \frac{1}{2}I_{4/5}\left(1,\frac{1}{2}\right)-\frac{1}{2}\sum_{k=1}^{n-1}\frac{\Gamma(\frac{2k+1}{2})}{\Gamma(k+1)\Gamma(\frac{1}{2})}\left(\frac{4}{5}\right)^{k}\left(\frac{1}{5}\right)^{\frac{1}{2}}\\ &= \frac{1}{2}-\frac{1}{2\sqrt{5}}-\frac{1}{2}\sum_{k=1}^{n-1}\binom{2k}{k}\left(\frac{1}{5}\right)^{k+\frac{1}{2}}\end{aligned}$$

Problem 318 (B. C. Hombas, Athens, Greece)
In proving the Central Limit Theorem the following equality was applied

$$\lim_{h\to 0}\int_{-\infty}^{\infty}\frac{\sin ht}{ht}\mathrm{e}^{-i(a+h)t}\phi(t)\,\mathrm{d}t=\int_{-\infty}^{\infty}\mathrm{e}^{-iat}\phi(t)\,\mathrm{d}t$$

provided that $\int_{-\infty}^{\infty}|\phi(t)|\,\mathrm{d}t=k<+\infty$. Prove this result.

The problem has been solved by the proposer. The result in fact follows immediately by Lebesgue's dominated convergence theorem.

Problem 319 (A. A. Jagers, W. C. M. Kallenberg)
Let X and Y be i.i.d. random variables. It is obvious that $X+Y$ and $X-Y$ have the same spread in the sense that $\mathrm{Var}(X+Y)=\mathrm{Var}(X-Y)$. In the L_1-sense this is not the case. Show that $\mathrm{E}|X+Y|\geq\mathrm{E}|X-Y|$ with equality iff X is symmetric, i.e. the distributions of $c-X$ and $X-c$ are the same, for some constant $c\in\mathbb{R}$.

In the previous formulation of the problem the name of A. A. Jagers was erroneously omitted. As L. A. Klein Haneveld remarked, $c=0$ should be taken, otherwise $X=Y=1$ almost surely would be a counterexample. In case $c=0$, problem 314 is obtained. A reformulation of the problem is given in the section "New Problems".

*Problem 320*** (R. D. Gill)
Let F and G be continuous distribution functions. Their QQ-plot is the graph of the quantiles $(x_p, y_p)=(F^{-1}(p), G^{-1}(p)), 0<p<1$, i.e. a graph of the function $y=G^{-1}(F(x))$. This monotone function transforms the distribution of $X\sim F$ into $Y=F^{-1}(G(X))$. If the QQ-plot is convex one says that Y is more skewed to the right than X. Suppose now X and Y are gamma distributed with parameters (α, r) and (β, r). Show that the gamma distribution with smaller shape parameter is more skew to the right.

Remark. *In his Ph.D. thesis W. R. van Zwet mentions that he managed to prove this after many pages of analysis. The pages are in a box in the archive of the now demolished former location of the Mathematical Center (Boerhaavestraat, Amsterdam), so presumably now lost forever!*

(Reference: W. R. van Zwet (1964), Convex transformations of random variables; MC Tract 7, Mathematical Centre, Amsterdam).

No solutions to this beautiful problem were submitted, it is kept open until further notice.

Problem 321 (A. H. Kroese)
Let $\text{med}(\chi^2_{p;\gamma})$ be the median of $\chi^2_{p;\gamma}$ the non-central chi-square distribution with p degrees of freedom and non-centrality parameter γ.

(I) Prove that there exists a constant $c > 0$ such that $\text{med}(\chi^2_{2;\gamma}) \geq \gamma + c$ for all $\gamma \geq 0$.

(I**) Numerical computations suggest that $c = 1$ is the largest value such that $\text{med}(\chi^2_{2;\gamma}) \geq \gamma + c$ for all $\gamma \geq 0$. Is this suggestion true and can it be proved?

The proposer has solved part I. No other solutions were received. Also this problem will be kept open.

Solution of part I by the proposer
Let Z_1, Z_2 be i.i.d. $\mathcal{N}(0,1)$. Then for all $\gamma > 0$,

$$(Z_1 + \sqrt{\gamma})^2 + Z_2^2 \sim \chi^2_{2,\gamma}$$

We have to prove that there exists a constant $c > 0$ such that $P((Z_1 + \sqrt{\gamma})^2 + Z_2^2 \leq \gamma + c) \leq ½$, for all $\gamma \geq 0$.

$$\begin{aligned}
P((Z_1 + \sqrt{\gamma})^2 + Z_2^2 \leq \gamma + c) &= P(Z_1^2 + Z_2^2 + 2\sqrt{\gamma}Z_1 \leq c) \leq P(Z_2^2 + 2\sqrt{\gamma}Z_1 \leq c) \\
&\leq P\left(Z_1 \leq -\frac{c}{2\sqrt{\gamma}}\right) + P\left(-\frac{c}{2\sqrt{\gamma}} \leq Z_1 \leq 0\right).P(Z_2^2 \leq 2c) \\
&\quad + P\left(0 \leq Z_1 \leq \frac{c}{2\sqrt{\gamma}}\right).P(Z_2^2 \leq c) \\
&= P\left(Z_1 \leq -\frac{c}{2\sqrt{\gamma}}\right) + P\left(-\frac{c}{2\sqrt{\gamma}} \leq Z_1 \leq 0\right)[P(Z_2^2 \leq 2c) + P(Z_2^2 \leq c)] \qquad (1)
\end{aligned}$$

Now choose c such that $P(Z_2^2 \leq 2c) + P(Z_2^2 \leq c) \leq 1$. ($c$ has to be smaller than approximately 0.32.) Then (1) is smaller than or equal to $P(Z_1 \leq 0) = ½$.

Problem 322 (I. Adan and J. Resing)

a. Let X be a random variable on the integers with mean $\mu (0 < \mu \leq 1)$. Prove that

$$\text{Var } X \geq \text{Var } Y$$

with equality if and only if $X = Y$, where Y is the random variable on 0 and 1 with probability distribution

$$P(Y = 1) = 1 - P(Y = 0) = \mu$$

b. Let (μ, γ) be a pair of real numbers with $\mu > 0$ and $\gamma \geq 0$. Prove that there exists a random variable X on the non-negative integers with mean μ and coefficient of variation γ if and only if

$$\gamma^2 \geq \frac{2k+1}{\mu} - \frac{k(k+1)}{\mu^2} - 1$$

where k is the unique integer satisfying $k < \mu \leq k+1$.

Note: For each pair (μ, γ) with $\mu > 0$ and $\gamma \geq 0$ there exists a continuous random variable $X \geq 0$ with mean μ and coefficient of variation γ. Part (b) of the problem states that this no longer true for discrete random variables (see Fig. 1).

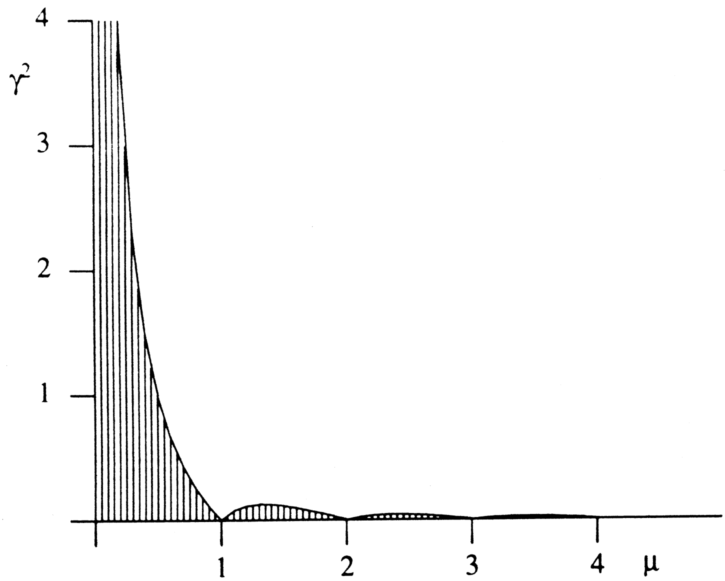

Fig. 1 For points (μ, γ^2) in the shaded area there exists no discrete random variable $X \geq 0$ with mean μ and squared coefficient of variation γ^2.

The problem has been solved by A. A. JAGERS, G. JONGLOED (Part (a)) and THE PROPOSERS. We publish two solutions.

Solution by THE PROPOSERS.

a. Since X is concentrated on the integers, it follows that

$$EX^2 \geq EX$$

with equality if and only if $X = Y$. Hence

$$\text{Var } X = EX^2 - (EX)^2 \geq EX - (EX)^2 = \mu(1-\mu) = \text{Var } Y$$

with equality if and only if $X = Y$.

b. For $n = 1, 2, \ldots$ let X_n be a random variable on k and $k+n$ with probability distribution

$$P(X_n = k + n) = 1 - P(X_n = k) = \frac{\mu - k}{n}$$

then it is easily seen that the mean $\mathrm{E}X_n$ and squared coefficient of variation c_n^2 of X_n are given by

$$\mathrm{E}X_n = \mu, \quad c_n^2 = \frac{2k+n}{\mu} - \frac{k(k+n)}{\mu^2} - 1$$

From Part (a) it follows for any random variable X on the integers with mean $\mu > 0$ that

$$c_X^2 \geq c_1^2$$

Hence the inequality in Part (b) is a necessary condition. To prove that it is also a sufficient condition, first note that c_n^2 is strictly increasing in n, and $c_n^2 \uparrow \infty$ as $n \to \infty$. So, for γ^2 satisfying the inequality in (b) there exists an n with

$$c_n^2 \leq \gamma^2 \leq c_{n+1}^2$$

and it is readily verified that the random variable X defined by

$$X = \begin{cases} X_{n+1} & \text{w.p. } \dfrac{\gamma^2 - c_n^2}{c_{n+1}^2 - c_n^2} \\ X_n & \text{w.p. } \dfrac{c_{n+1}^2 - \gamma^2}{c_{n+1}^2 - c_n^2} \end{cases}$$

has mean μ and squared coefficient of variation γ^2.

Solution by A. A. Jagers

Let $\mathcal{C}$ be the class of all integer-valued random variables with mean μ. Define $k \in \mathbb{Z}$ by $k < \mu \leq k+1$ and assume (the opposite case being similar) that $k + 1 - \mu \leq \mu - k$.

By weak compactness the minimum value of Var X over $X \in \mathcal{C}$ is attained, for a variable Z, say. We claim that the distribution of Z is concentrated on $\{k, k+1\}$.

In fact, if $P\{Z \geq k+2\}$ is positive then so is $P\{Z \geq k\}$, since $\mu \leq k+1$; and, if $0 < p \leq \min(P\{Z \geq k+2\}, P(Z \leq k))$, then by moving a probability mass p from $\{k+2, k+3, \ldots\}$ one unit to the left and the same amount p from $\{\ldots, k-1, k\}$ one unit to the right we would obtain a new distribution on $\mathbb{Z}$ (from the distribution of Z) with the same mean but lesser variance (the condition $k+1-\mu \leq \mu - k$ comes into use here).

Since additionally $\mathrm{E}Z = \mu$, it follows that $P\{Z = k\} = k + 1 - \mu$, $P\{Z = k+1\} = \mu - k$ and Var $Z = (k+1-\mu)(\mu - k)$, the minimum value in question. If $k = 0$, then

Z has the same distribution as the variable Y from Part (a). As for Part (b), for $X \in \mathcal{C}$ we have

$$\gamma^2 = \frac{\text{Var } X}{\mu^2} \geq \frac{\text{Var } Z}{\mu^2} = \frac{2k+1}{\mu} - \frac{k(k+1)}{\mu^2} - 1$$

as required. Conversely, if the distribution of $X \in \mathcal{C}$ is given by $P\{X = k\} = k + 1 - \mu + t, P\{X = k+1\} = \mu - k - nt/(n-1) \geq 0$ and $P\{X = k+n\} = t/(n-1)$ with $k \geq 0, n \geq 2$ and $t > 0$, then $\text{E}X = \mu$ and $\text{Var } X = \text{Var } Z + nt$; and nt takes on all positive real values, if n runs through $\{2, 3, \ldots\}$ and t varies between 0 and $(\mu - k)/2$.

New Problems

Problem 319 (A. A. Jagers, W. C. M. Kallenberg), revision
Let X and Y be i.i.d. random variables. It is obvious that $X + Y$ and $X - Y$ have the same spread in the sense that $\text{Var}(X + Y) = \text{Var}(X - Y)$. In the L_1-sense this is not the case. From problem 314 it follows that

$$\text{E}|X + Y - \text{E}(X + Y)| \geq \text{E}|X - Y - \text{E}(X - Y)|$$

with equality iff $X - \text{E}X$ is symmetric around 0. Here the expectation is used as a measure of location, which is quite natural for spread in the L_2-sense, the variance. In the L_1-sense the median is more natural ($\text{E}|Z - a|$ is minimized by $a = \text{med } Z$).

Show that

$$\text{E}|X + Y - \text{med}(X + Y)| \geq \text{E}|X - Y - \text{med}(X - Y)|$$

When does the equality sign hold?

Problem 326 (R. Jans, Louvain, Belgium)
In a lottery game, n numbers are randomly drawn from a set of m numbers. On a ticket you may fill out n numbers hoping that they will match the n numbers selected. We want to know the minimum number of tickets one has to fill out in order to be sure that among the tickets filled out there is at least one ticket that has $n - 1$ or (or $n - 2, n - 3, \ldots$) numbers correct. Can this problem be solved using calculus of combinations?

Problem 327 (K. van Harn and F. W. Steutel)
For $n \in \mathbb{N}$, let f_n be the gamma probability density with shape parameter $t = n/2$ and scale parameter $\lambda = 1$, so

$$f_n(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} x^{(n/2)-1} \text{e}^{-x} \quad [x > 0]$$

Define the function: f: $(0, \infty) \to \mathbb{R}$ by

$$f(x) := 2\sum_{n=1}^{\infty}(-1)^{n-1}f_n(x) \quad [x > 0]$$

Show that f is a probability density.

*Problem 328*** (R. THEODORESCU, STE. FOY, CANADA)
Let $m_n(t)$ be the nth moment of a Poisson distribution with parameter $t > 0$. It is known that

$$m_n(t) = \sum_{j=0}^{n} S(n,j)t^j$$

where $S(n,j)$ is the Stirling number of the second kind; moreover,

$$m_{n+1}(t) = t(m_n(t) + (\mathrm{d}/\mathrm{d}t)m_n(t))$$

Consider the normalized moments $c_n(t) = m_n(t)/n!$.

CONJECTURE: *for each* $t > 0$

$$c_{n+1}^2(t) > c_n(t)c_{n+2}(t), \quad n \geq 1 \qquad (*)$$

For $n = 0$ $(*)$ *holds for* $t \geq 1$. *Is the logconcavity Conjecture* $(*)$ *true?*

Problems marked with $*$ are nonelementary, of problems marked with $**$ no solution is known to the editor; unmarked problems are not necessarily simple. Solutions of the problems in this issue should arrive before September 30, 1996. Problems (preferably with solutions) and solutions (type-written on separate sheets baring the name of the solver) are welcomed by the column editor.