The Annals of Probability
1982, Vol. 10, No. 3, 799–809

ON UPPER AND LOWER BOUNDS FOR THE VARIANCE OF A FUNCTION OF A RANDOM VARIABLE[1]

BY THEOPHILOS CACOULLOS

University of Athens

Chernoff (1981) obtained an upper bound for the variance of a function of a standard normal random variable, using Hermite polynomials. Chen (1980) gave a different proof, using the Cauchy-Schwarz inequality, and extended the inequality to the case of a multivariate normal. Here it is shown how similar upper bounds can be obtained for other distributions, including discrete ones. Moreover, by using a variation of the Cramér-Rao inequality, analogous lower bounds are given for the variance of a function of a random variable which satisfies the usual regularity conditions. Matrix inequalities are also obtained. All these bounds involve the first two moments of derivatives or differences of the function.

1. Introduction. Let X have the standard normal distribution $N(0, 1)$. Chernoff (1981), using Hermite polynomials, proved the inequality

$$\operatorname{Var}[g(X)] \le E[g'(X)]^2, \tag{1.1}$$

if g is an absolutely continuous real-valued function and $g(X)$ has finite variance; equality holds if and only if $g(x) = ax + b$ for some constants a and b.

Chen (1980) proved a multivariate extension of (1.1) using the Cauchy-Schwarz inequality. Specifically, if $X_1, \cdots, X_k$ are independent identically distributed (i.i.d.) normal $N(0, 1)$ r.v.'s and $g, g_1, \cdots, g_k$ real-valued Borel measurable functions defined on R^k such that

$$g(x_1, \cdots, x_k) = \int_0^{x_i} g_i(x_1, \cdots, x_{i-1}, t, x_{i+1}, \cdots, x_k)\, dt + g(x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_k)$$

for $i = 1, \cdots, k$, then

$$\operatorname{Var}[g(X_1, \cdots, X_k)] \le \textstyle\sum_{i=1}^k E[g_i(X_1, \cdots, X_k)]^2; \tag{1.2}$$

as in (1.1), equality holds if and only if $g(x_1, \cdots, x_k)$ is a linear function of $x_1, \cdots, x_k$.

These inequalities are relevant to solving variations of the classical isoperimetric problem, which in turn is related to the problem of data compression in the theory of element identification; see Chernoff (1980).

The preceding results motivated the search for similar variance bounds of functions of other random variables, including discrete ones. Moreover, interestingly enough, using the well-known Cramér-Rao (C-R) inequality—another by-product of the Cauchy-Schwarz inequality—a lower bound was obtained for the variance of a function $g(X)$ of a random variable X which fulfills the regularity conditions stated in Lemma 2.2 (Section 2). Finally, a matrix extension of (1.2) is given along with a lower bound, obtained from the C-R inequality for multiparameter distributions. A common feature of all these variance bounds is that they involve the first two moments of derivatives of g in the continuous case and differences in the discrete case.

2. Preliminary general results. For our purposes, we prove two lemmas, of some

Received April 1981; revised October 1981.
[1] Supported in part by NATO/HEINEMANN Senior Fellowship No. 146.
AMS 1980 *subject classification.* Primary 60E05; secondary 62F10, 62H99.
Key words and phrases. Variance bounds, Cramér-Rao inequality.

interest by themselves. The proof of the following lemma follows the lines of Lemma 2.2 of Chen (1980).

LEMMA 2.1. *Let X be a continuous* r.v. *with density function $f(x)$. Let g and g' be real-valued functions on R such that g is an indefinite intergral of g', and* $\mathrm{Var}[g(X)] < \infty$. *Then*

$$\text{(2.1)} \qquad \mathrm{Var}[g(X)] \leq \int_0^\infty \int_t^\infty x f(x)[g'(t)]^2\, dx\, dt - \int_{-\infty}^0 \int_{-\infty}^t x f(x)[g'(t)]^2\, dx\, dt.$$

PROOF. Since $g(x)$ and $\int_0^x g'(t)\, dt$ differ by a constant, we have

$$\mathrm{Var}[g(X)] = \mathrm{Var}\left[\int_0^X g'(t)\, dt\right] \leq E\left[\int_0^X g'(t)\, dt\right]^2.$$

Now by the Cauchy-Schwarz inequality

$$\mathrm{Var}[g(X)] \leq E\left[\int_0^X 1^2\, dt \int_0^X [g'(t)]^2\, dt\right] = E\left\{X \int_0^X [g'(t)]^2\, dt\right\}$$

$$= \int_{-\infty}^\infty x f(x) \int_0^x [g'(t)]^2\, dt\, dx = \int_0^\infty \int_0^x x f(x)[g'(t)]^2\, dt\, dx$$

$$+ \int_{-\infty}^0 \int_0^x x f(x)[g'(t)]^2\, dt\, dx = \int_0^\infty \int_t^\infty x f(x)[g'(t)]^2\, dx\, dt$$

$$- \int_{-\infty}^0 \int_{-\infty}^t x f(x)[g'(t)]^2\, dx\, dt,$$

which proves the lemma.

The next lemma is an immediate corollary of the derivation of the Cramér-Rao inequality.

LEMMA 2.2. *Let $X_1, \cdots, X_n$ be independent* r.v.'s *with common density $f(\cdot, \theta)$, $\theta \in \Theta$, where Θ is an open interval in R. Suppose f satisfies the regularity conditions*:

(i) *$f(x, \theta)$ is positive on a set $\mathcal{S}$ independent of θ*
(ii) *For each $\theta \in \Theta$*

$$E_\theta\left[\frac{\partial}{\partial \theta} \log f(X, \theta)\right] = 0.$$

Let $g: R^n \to R$ a function such that

$$E_\theta\left[g(\mathbf{X}) \frac{\partial}{\partial \theta} \log \textstyle\prod_{i=1}^n f(X_i, \theta)\right] \equiv a(\theta)$$

exists; we set $\mathbf{X} = (X_1, \cdots, X_n)$. Then

$$\text{(2.2)} \qquad \mathrm{Var}[g(\mathbf{X})] \geq \frac{[a(\theta)]^2}{n J(\theta)}$$

where the Fisher information (number) $J(\theta)$ in X_i at θ is given by

$$J(\theta) = E\left[\frac{\partial}{\partial \theta} \log f(X_i, \theta)\right]^2.$$

LEMMA 2.3 *Let X be a non-negative integer-valued* r.v. *with probability function $p(x)$, and $g(x)$ a real-valued function defined on* $\{0, 1, 2, \cdots\}$ *such that the* $\mathrm{Var}[g(X)] < \infty$. *Then*

$$\mathrm{Var}[g(X)] \le \textstyle\sum_{k=0}^{\infty} [\Delta g(k)]^2 \sum_{x=k+1}^{\infty} xp(x) \tag{2.3}$$

where Δ denotes the forward difference operator, i.e., $\Delta g(x) = g(x+1) - g(x)$.

PROOF. We write $g(x)$ in the form

$$g(x) = \textstyle\sum_{k=0}^{x-1} \Delta g(k) + g(0),$$

so that by the same argument of Lemma 2.1

$$\begin{aligned}\mathrm{Var}(g(X)) &= \mathrm{Var}(\textstyle\sum_{k=0}^{X-1} \Delta g(k)) \le E\{\sum_{k=0}^{X-1} \Delta g(k)\}^2 \\ &\le E\{\textstyle\sum_{k=0}^{X-1} 1^2 \sum_{k=0}^{X-1} [\Delta g(k)]^2\} = E\{X \sum_{k=0}^{X-1} [\Delta g(k)]^2\} \\ &= \textstyle\sum_{x=1}^{\infty} x\{\sum_{k=0}^{x-1} [\Delta g(k)]^2 p(x)\} = \sum_{k=0}^{\infty} [\Delta g(k)]^2 \sum_{x=k+1}^{\infty} xp(x).\end{aligned}$$

Finally, we give the analogue of (2.2) for the multiparameter case (cf. Rao (1973), page 326).

LEMMA 2.4. *Let X be a random variable with probability density $f(x, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)$. Let $\mathbf{g}(x) = (g_1(x), \cdots, g_r(x))$ such that the* $\mathrm{Var}[g_i(X)] < \infty$, $i = 1, \cdots, r$. *Then the dispersion matrix $D[\mathbf{g}(X)] = (\sigma_{ij})$, $\sigma_{ij} = \mathrm{Cov}(g_i(X), g_j(X))$, satisfies*

$$D[\mathbf{g}(X)] \ge \Lambda J^{-1}(\boldsymbol{\theta})\Lambda' \tag{2.4}$$

where J^{-1} is the inverse of the information matrix $J(\theta) = (J_{ij})$ with

$$J_{ij} = E\left\{-\frac{\partial^2 \log f(x, \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right\}, \quad i, j = 1, \cdots, k$$

and $\Lambda = (\lambda_{ij})$ with

$$\lambda_{ij} = \int_{-\infty}^{\infty} g_i(x) \frac{\partial f(x, \boldsymbol{\theta})}{\partial\theta_j} dx, \quad \begin{matrix} i = 1, \cdots, r \\ j = 1, \cdots, k. \end{matrix} \tag{2.5}$$

Note that X may also be a random vector.
Inequality (2.4) is to be understood in the sense of the following.

DEFINITION 2.1. If A and B are non-negative definite matrices, then $A \ge B$ if and only if $A - B$ is non-negative definite.

3. The normal distribution. In this section, we apply Lemmas 2.1 and 2.2 to obtain upper and lower bounds for the variance of a function of a standard normal random variable $N(0, 1)$.

PROPOSITION 3.1. *Let X follow the normal distribution $N(0, 1)$ and $g(x)$ be a function satisfying the conditions of Lemma* 2.1 *and* $E\{|g'(X)|\} < \infty$. *Then*

$$E^2[g'(X)] \le \mathrm{Var}[g(X)] \le E[g'(X)]^2, \tag{3.1}$$

where both equalities hold if and only if $g(x)$ is linear.

PROOF. The second inequality, i.e. Chernoff's inequality (1.1), follows immediately from (2.1) with $f(x) = \varphi(x) = (2\pi)^{-1/2}\exp\{-\frac{1}{2}x^2\}$ and the fact (cf. Lemma 2.2 of Chen, 1980)

$$\int_t^{\infty} x\varphi(x)\, dx = \varphi(t), \quad -\int_{-\infty}^{t} x\varphi(x)\, dx = \varphi(t).$$

For the first inequality, we apply (2.2) with $n = 1$ and $\theta = \mu$ for a r.v. X which has the normal distribution $N(\mu, 1)$. We have

$$\int_{-\infty}^{\infty} g(x) \frac{\partial}{\partial\mu} \frac{1}{\sqrt{2\pi}} \exp[-(x-\mu)^2/2]\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x)(x-\mu)\exp[-(x-\mu)^2/2]\, dx$$

$$= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x)\, d\exp[-(x-\mu)^2/2] \tag{3.2}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g'(x)\exp[-(x-\mu)^2/2]\, dx = E[g'(X)].$$

Here, in integrating by parts, we used the fact that for some x sequences tending to $+\infty$ or $-\infty$

$$\lim_{X\to\pm\infty} g(x)\exp[-(x-\mu)^2/2] = 0,$$

which follows from $E\{|g'(X)|\} < \infty$.

Since (3.2) and $J(\mu) = 1$ hold for every μ in R, the proof is complete. It is easily verified that the first equality in (3.1) holds if and only if $g(X) = aX + b$ for some constants a and b. Of course, this is in agreement with the fact that X is the efficient estimator of μ.

REMARK 3.1. If X is $N(\mu, \sigma^2)$, then $X = \sigma Z + \mu$ where Z is $N(0, 1)$; therefore, setting $g(x) = g_0(z)$ we have $g_0'(z) = \sigma g'(x)$. Thus applying (3.1) to the function $g_0(Z)$, we obtain the following.

PROPOSITION 3.2. *Let X be $N(\mu, \sigma^2)$ and $g(x)$ as above. Then*

$$\sigma^2 E^2(g'(X)] \le \operatorname{Var}[g(X)] \le \sigma^2 E[g'(X)]^2 \tag{3.3}$$

where equality holds if and only if $g(x) = ax + b$ for some constants a and b.

Now we apply (2.2) to the density of $N(0, \sigma^2)$ taking $\theta = \sigma^2$. This gives the following.

PROPOSITION 3.3. *Let X be $N(0, \sigma^2)$ and $g(x)$ and $g'(x)$ absolutely continuous such that $E\{|g''(X)|\} < \infty$. Then*

$$\operatorname{Var}[g(X)] \ge \frac{\sigma^4}{2} E^2[g''(X)], \tag{3.4}$$

where equality holds if and only if $g(x) = ax^2 + b$ for some constants a and b.

PROOF. Integrating by parts twice (cf. the proof of (3.1) and McShane, 1944, pages 209, 332), we have

$$\int_{-\infty}^{\infty} g(x) \frac{\partial}{\partial\theta} \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta}\, dx = \frac{1}{2\sqrt{2\pi\theta}} \int_{-\infty}^{\infty} g''(x) e^{-x^2/2\theta}\, dx = \frac{1}{2} E[g''(X)].$$

Since $J(\theta) = (2\theta^2)^{-1} = (2\sigma^4)^{-1}$, the desired inequality (3.4) readily follows from (2.2). Noting that equality holds if and only if

$$\frac{\partial}{\partial\theta} \log \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta} = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2} = cg(x)$$

completes the proof.

NOTE. The condition on $g(x)$ in order to attain the lower bound in (3.4) is a consequence of the fact that X^2 is the efficient unbiased estimator of σ^2, if X is $N(0, \sigma^2)$.

The following simple application of (2.4) to the $N(\mu, \sigma^2)$ throws more light on the nature of the lower variance bounds given by (3.3) and (3.4).

PROPOSITION 3.4. *Let X be $N(\mu, \sigma^2)$ and $g(x)$ as in Proposition 3.3 such that $E\{|g'(X)|\} < \infty$, $E\{|g''(X)|\} < \infty$. Then*

$$\mathrm{Var}[g(X)] \geq \sigma^2 E^2[g'(X)] + ½\sigma^4 E^2[g''(X)] \tag{3.5}$$

where equality holds if and only if $g(x) = ax^2 + bx + c$ for some constants a, b and c.

It is observed that (3.5) improves the lower bound provided by either (3.3) or (3.4). This is so, because the lower limit of (3.3) is applicable whether σ^2 is known or unknown (X is the unbiased efficient estimator of μ), whereas the lower bound of (3.4) is applicable only if μ is known. In general, if several parameters $\theta_1, \cdots, \theta_k$ are involved in an estimation problem, then the C-R lower bound for an estimate of θ_i as a single parameter may not be attained unless the estimate happens to be independent of the other parameters. Otherwise, as in the present case, the lower limit may increase (cf. Rao, 1973, page 327).

On the other hand, the lower bound of the variance of $g(X)$, regarded as an estimator of its mean $E[g(X)] = \gamma(\mu, \sigma^2)$, can be expressed in terms of the information matrix $J_X(\mu, \sigma^2)$ and the derivatives of γ, independently of any chosen unbiased estimator of $\gamma(\mu, \sigma^2)$, such as $g(X)$. Indeed, by the Cramér-Rao inequality for the multiparameter case, we find

$$\mathrm{Var}[g(X)] \geq \sigma^2\left(\frac{\partial\gamma}{\partial\mu}\right)^2 + 2\sigma^4\left(\frac{\partial\gamma}{\partial\sigma^2}\right)^2.$$

However, this lower bound involves the first two derivatives of g, as shown by (3.5), reflecting the dependence on g.

Let us now consider bounds for the variance of a function $g(X_1, \cdots, X_n)$ where the X_i are normal. Using (2.2) with $\theta = \mu$ and proceeding as in Proposition 3.1, we obtain the following.

PROPOSITION 3.5. *Let $\mathbf{X} = (X_1, \cdots, X_n)$ where the X_i are independent $N(\mu, \sigma^2)$. Suppose $g(x_1, \cdots, x_n)$ has partial derivatives $g_i(x_1, \cdots, x_n) = (\partial/\partial x_i)g(x_1, \cdots, x_n)$ and $E\{|g_i(\mathbf{X})|\} < \infty$, $i = 1, \cdots, n$. Then*

$$\mathrm{Var}[g(\mathbf{X})] \geq \frac{\sigma^2}{n}\{\textstyle\sum_{i=1}^n E[g_i(\mathbf{X})]\}^2 \tag{3.6}$$

where equality holds if and only if $g(\mathbf{X}) = a\sum_{i=1}^n X_i + b$ for some constants a and b.

Combining (3.6) with the extension of (1.2) by (3.3) and (3.8), we have

$$\frac{\sigma^2}{n}\{\textstyle\sum_{i=1}^n Eg_i(\mathbf{X})\}^2 \leq \mathrm{Var}[g(\mathbf{X})] \leq \sigma^2 \textstyle\sum^n ss_{i=1} E[g_i(\mathbf{X})]^2.$$

Applying (2.2) again, with $\theta = \sigma^2$ and proceeding as in Proposition 3.3, we have the following.

PROPOSITION 3.6. *Let $\mathbf{X} = (X_1, \cdots, X_n)$ where the X_i are independent $N(0, \sigma^2)$. Let $g(x)$ be absolutely continuous with $g_i(x) = (\partial/\partial x_i)g(x)$ absolutely continuous and suppose $E\{|g_{ii}(X)|\} < \infty$, where $g_{ii}(x) = (\partial^2/\partial x_i^2)g(x)$, $i = 1, \cdots, n$. Then*

$$\mathrm{Var}[g(\mathbf{X})] \geq \frac{\sigma^4}{2n} E^2\{\textstyle\sum_{i=1}^n g_{ii}(\mathbf{X})\}$$

where equality holds if and only if $g(\mathbf{X}) = a\sum_{i=1}^n X_i^2 + b$ for some constants a and b.

As an example of an application of (2.4), we consider the case of a multivariate normal $N(\boldsymbol{\mu}, \Sigma)$, taking $\boldsymbol{\theta} = \boldsymbol{\mu} = (\mu_1, \cdots, \mu_k)$ and $r = 1$. We will show the following.

PROPOSITION 3.7. *Let* $\mathbf{X}$ *be* $N(\boldsymbol{\mu}, \Sigma)$ *and* $g(\mathbf{x})$ *a differentiable real-valued function defined on* R^k, *with gradient* $\nabla g = (g_1, \cdots, g_k)^t$ *where* $g_i = D_i g$ *is the partial derivative of* g *with respect to* x_i *and* t *denotes transpose. Suppose* $E\{|g_i(X)|\} < \infty$. *Then*

$$\mathrm{Var}[g(\mathbf{X})] \geq E[\nabla^t g(\mathbf{X})] \Sigma\, E[\nabla g(\mathbf{X})] \tag{3.7}$$

where equality holds if and only if $g(x) = a_1 x_1 + \cdots + a_k x_k + b$ *for some constants* $a_1, \cdots, a_k$ *and* b.

PROOF. This is straightforward; it is easily verified that the information matrix, using $\boldsymbol{\theta} = \boldsymbol{\mu}$

$$J = \left(-E\left[\frac{\partial^2 \log f(\mathbf{x}, \boldsymbol{\mu})}{\partial\mu_i \partial\mu_j}\right]\right) = \Sigma^{-1},$$

whereas the vector $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_k)^t$ (cf(2.5)) is given by

$$\lambda_j = \int_{R^k} g(\mathbf{x}) \frac{\partial f(\mathbf{x}, \boldsymbol{\mu})}{\partial \mu_j}\, d\mathbf{x} = -\int_{R^k} g(\mathbf{x}) D_j f(\mathbf{x}, \boldsymbol{\mu})\, d\mathbf{x}$$

$$= \int_{R^k} [D_j g(\mathbf{x})] f(\mathbf{x}, \boldsymbol{\mu})\, d\mathbf{x} = E[g_j(\mathbf{X})].$$

Hence from (2.4), we obtain (3.7).

At this point it should be mentioned that Chen (1980) obtained the following upper bound inequality for the case $N(\mathbf{0}, \Sigma)$:

$$\mathrm{Var}[g(\mathbf{X})] \leq E[\nabla^t g(\mathbf{X}) \Sigma \nabla g(\mathbf{X})], \tag{3.8}$$

which is a generalization of (1.2) when g has a differential at every point in R^k. Combining this inequality with (3.7), we obtain the following.

COROLLARY 3.1. *Let* $\mathbf{X}$ *be* $N(\mathbf{0}, \Sigma)$ *and* g *have a differential at every point in* R^k. *Then*

$$[E\, \nabla^t g(\mathbf{X})] \Sigma [E\, \nabla g(\mathbf{X})] \leq \mathrm{Var}[g(\mathbf{X})] \leq E[\nabla^t g(\mathbf{X}) \Sigma \nabla g(\mathbf{X})] \tag{3.9}$$

where the two bounds coincide if and only if $g(\mathbf{x})$ *is linear, as in* (3.7)

We close this section with a matrix analogue of (3.9).

PROPOSITION 3.8. *Let* $\mathbf{X}$ *be* $N(0, I)$ *and let* $\mathbf{g}(\mathbf{X}) = (g_1(\mathbf{X}), \cdots, g_r(\mathbf{X}))$ *where each* g_i *is differentiable in* R^k. *Set*

$$g_{ij} = \frac{\partial g_i(\mathbf{x})}{\partial x_j}, \quad \Gamma = (g_{ij}), \quad i = 1, \cdots, r, \quad j = 1, \cdots, k.$$

Then (*see Definition* 2.1)

$$(E\Gamma)(E\Gamma)^t \leq D[\mathbf{g}(\mathbf{X})] \leq E(\Gamma\Gamma^t) \tag{3.10}$$

where the two bounds coincide if and only if each g_i *is linear, that is,*

$$\mathbf{g}(\mathbf{X}) = A\,\mathbf{X} + \mathbf{b} \tag{3.11}$$

where A *is a matrix of constants and* $\mathbf{b}$ *is a constant vector.*

PROOF. The lower bound in (3.10) is obtained by using (2.4) for the normal $N(\boldsymbol{\mu}, I)$

and the vector function $\mathbf{g}$ and proceeding as in Proposition 3.5. Thus, from (2.5) and (3.9), we have

$$\lambda_{ij} = E[g_{ij}(\mathbf{X})], \quad i = 1, \cdots, r, \quad j = 1, \cdots, k$$

and since the information matrix J is now the identity matrix, we conclude the first inequality in (3.10).

For the second inequality, we use (1.2). It suffices to show that for any constant vector $\mathbf{c} = (c_1, \cdots, c_r)^t$

$$\operatorname{Var}[\mathbf{c}^t\mathbf{g}(\mathbf{X})] = \mathbf{c}^t D[\mathbf{g}(\mathbf{X})]c \le \mathbf{c}^t E(\Gamma\Gamma^t)\mathbf{c}.$$

This however is a consequence of the fact that by (1.2) or (3.9)

$$\operatorname{Var}[\mathbf{c}^t\mathbf{g}(\mathbf{X})] \le E\{\nabla^t[\mathbf{c}^t\mathbf{g}(\mathbf{X})]\nabla[\mathbf{c}^t\mathbf{g}(\mathbf{X})]\} = E[(\mathbf{c}^t\Gamma)(\Gamma^t\mathbf{c})] = \mathbf{c}^t E(\Gamma\Gamma^t)\mathbf{c}.$$

Thus and by Corollary 3.1, as regards the equality of the two bounds, our proof is complete.

NOTE. The required modification of (3.10) when X has a multivariate normal with mean zero and general covariance matrix Σ is obvious.

4. The exponential distribution. Here we apply the general inequalities (2.1) and (2.2) to the exponential density

$$f(x, \theta) = \theta e^{-\theta x}, \quad x > 0. \tag{4.1}$$

PROPOSITION 4.1. *Let X have the exponential density* (4.1) *and let $g(x)$ be a differentiable function such that $g(X)$ has finite variance. Then*

$$\operatorname{Var}[g(X)] \le E\left[\int_0^X g'(t)\,dt\right]^2 \le \frac{1}{\theta^2} E\{[g'(X)]^2(1 + \theta X)\} \tag{4.2}$$

where both equalities hold if and only if $g(x)$ = constant.

PROOF. The left-hand inequality is trivial. The right-hand-side inequality follows from (2.1) noting that

$$\int_t^\infty \theta x e^{-\theta x}\,dx = \frac{1}{\theta}(\theta t + 1)e^{-\theta t}.$$

The first equality in (4.2) requires $E[\int_0^X g'(t)\,dt] = 0$, and the second equality requires that g be linear. Hence both equalities hold if and only if $g(x)$ is a constant.

Inequality (4.2) can be improved by noting that under (4.1)

$$E[g(X)] = \frac{1}{\theta} E[g'(X)] + g(0).$$

Hence, $E[\int_0^x g'(t)\,dt] = \frac{1}{\theta} E[g'(X)]$ and (4.2) gives

$$\operatorname{Var}(g(X)) = \operatorname{Var}\left[\int_0^X g'(t)\,dt\right] = E\left[\int_0^X g'(t)\,dt\right]^2 - E^2\left[\int_0^X g'(t)\,dt\right]$$
$$\le \frac{1}{\theta^2} E\{[g'(X)]^2 - E^2[g'(X)]\} + \frac{1}{\theta} E\{X[g'(X)]^2\}.$$

Note also that here equality holds if and only if the second equality in (4.2) holds, i.e., $g(x)$ is linear. Thus we have shown the following.

PROPOSITION 4.2. *Let X have the exponential density* (4.1) *and $g(x)$ be differentiable such that* $\mathrm{Var}[g(X)] < \infty$.

$$\mathrm{Var}[g(X)] \le \frac{1}{\theta^2}\,\mathrm{Var}[g'(X)] + \frac{1}{\theta}\,E\{X[g'(X)]^2\} \tag{4.3}$$

where equality hold if and only if $g(x)$ is linear.

As regards a lower bound for the variance of $g(X)$, we apply (2.2) for $n = 1$ to obtain the following.

PROPOSITION 4.3. *Under the assumptions of Proposition* 4.1

$$\mathrm{Var}[g(X)] \ge E^2[Xg'(X)] \tag{4.4}$$

where equality holds if and only if $g(x)$ is linear.

It may be added that the treatment of the gamma distribution is similar. A straightforward application of (2.2) with $n = 1$ and θ equal to the scale parameter λ yields the following.

PROPOSITION 4.4 *If $g(x)$ satisfies the assumptions of Proposition* 4.1 *and X has the density*

$$f(x, \lambda) = \frac{\lambda^k}{\Gamma(k)}\,x^{k-1}e^{-\lambda x}, \quad (x > 0,\ k > 0)$$

then

$$\mathrm{Var}[g(X)] \ge \frac{1}{k}\,E^2[Xg'(X)],$$

where equality holds if and only if $g(x)$ is linear.

We can easily obtain the analogue of (4.4) for the case of $\mathbf{X} = (X_1, \cdots, X_n)$ where the X_i are i.i.d. exponential r.v.'s with density given by (4.1). We have

$$J_{\mathbf{X}}(\theta) = nJ_{X_1}(\theta) = \frac{n}{\theta^2}$$

and if g is differentiable (in the positive orthant of R^n) with gradient $\nabla g = (g_1, \cdots, g_n)^t$, then

$$E\left[g(\mathbf{X})\,\frac{\partial}{\partial\theta}\log f(\mathbf{X}, \theta)\right] = \sum_{i=1}^n E\left[g(\mathbf{X})\,\frac{\partial}{\partial\theta}\log f(X_i, \theta)\right]$$

$$= -\frac{1}{\theta}\sum_{i=1}^n E[X_i g_i(\mathbf{X})].$$

Thus we have shown the inequality (cf.(3.6))

$$\frac{1}{n}\,E^2[\mathbf{X}\,\nabla g(\mathbf{X})] \le \mathrm{Var}[g(\mathbf{X})], \tag{4.5}$$

(treating $\mathbf{X}$ as a row vector).

Similarly, assuming different parameters for the X_i and using (2.4), we obtain the lower-bound inequality (cf.(3.7) when $\Sigma = I$)

$$\sum_{i=1}^n E^2[X_i g_i(\mathbf{X})] \le \mathrm{Var}[g(\mathbf{X})]$$

where equality holds if and only if g is linear in $X_1, \cdots, X_n$.

5. The Poisson distribution. The derivation of upper bounds for the variance of a function of a discrete r.v. is based on Lemma 2.3, whereas lower bounds are obtained from Lemma 2.2, as in the continuous case.

The main result for the Poisson distribution indicates that upper bounds do not have the simple form of lower bounds (cf. (3.1)).

PROPOSITION 5.1. *Let X be a Poisson* r.v. *with parameter λ and $g(x)$ a real-valued function defined on the set $N_0 = \{0, 1, 2, \cdots\}$ such that* $\operatorname{Var}[g(X)]$ *is finite. Then*

$$\lambda E^2[\Delta g(X)] \le \operatorname{Var}[g(X)] \le \lambda E[\Delta g(X)]^2 + \lambda \int_0^\lambda \{E_y[\Delta g(X)]^2\}\, dy \tag{5.1}$$

where E_y indicates expectation under a Poisson with parameter y and both equalities hold if $g(x)$ is a constant.

PROOF. For the first inequality, we apply (2.2) to the Poisson probability function $f(x, \theta)$, with $\theta = \lambda$, thus obtaining

$$\begin{aligned}\sum_{x=0}^\infty \left[g(x) \frac{\partial}{\partial\lambda} f(x, \lambda) \right] &= e^{-\lambda}\left[\sum_{x=1}^\infty \frac{\lambda^{x-1}}{(x-1)!} g(x) - \sum_{x=0}^\infty \frac{\lambda^x}{x!} g(x) \right] \\ &= e^{-\lambda} \sum_{x=0}^\infty \frac{\lambda^x}{x!} [g(x+1) - g(x)] = E[\Delta g(X)].\end{aligned}$$

Hence and from the fact that $J_X(\lambda) = 1/\lambda$, we obtain the lower bound in (5.1), which is attained only if g is linear.

For the upper bound, we write

$$g(x) = \sum_{k=0}^{x-1} \Delta g(k) + g(0)$$

and, by Lemma 2.3, setting $a_k \equiv \Delta g(k)$, we have

$$\operatorname{Var}[g(X)] \le e^{-\lambda} \sum_{k=0}^\infty a_k^2 \sum_{x \ge k+1} \frac{\lambda^x}{(x-1)!}. \tag{5.2}$$

To evaluate the second sum, we differentiate with respect to λ the well-known relation

$$\sum_{x \ge k+1} \frac{\lambda^x}{x!} = e^\lambda \int_0^\lambda \frac{1}{k!} e^{-y} y^k\, dy.$$

This yields

$$\sum_{x=k+1}^\infty x \frac{\lambda^x}{x!} = \lambda e^\lambda \left\{ \int_0^\lambda \frac{1}{k!} e^{-y} y^k\, dy + \frac{\lambda^k}{k!} e^{-\lambda} \right\}.$$

Hence the right-hand side of (5.2) is equal to

$$\lambda \int_0^\lambda \left(\sum_{k \ge 0} a_k^2 \frac{1}{k!} e^{-y} y^k \right) dy + \lambda \sum_{k \ge 0} a_k^2 e^{-\lambda} \frac{\lambda^k}{k!},$$

from which (5.1) follows. Therefore the proof of (5.1) is complete.

As regards a lower bound for the variance of $g(\mathbf{X})$ when the components $X_1, \cdots, X_n$ of $\mathbf{X}$ are independent Poisson r.v.'s with the same parameter λ, we obtain the inequality (cf.(3.6), (4.5))

$$\frac{\lambda}{n} E^2\left[\sum_{k=1}^n \Delta_k g(\mathbf{X})\right] \le \operatorname{Var}[g(\mathbf{X})]$$

where Δ_k indicates the forward difference operator operating on x_k, i.e.,

$$\Delta_k g(\mathbf{x}) = g(x_1, \cdots, x_k + 1, x_{k+1}, \cdots, x_n) - \Delta(x_1, \cdots, x_n).$$

We give also the lower variance bound which is obtained by using (2.4) when the X_i are independent with different parameters λ_ι. We find

$$\sum_{k=1}^{n} \lambda_\iota E^2[\Delta_k g(\mathbf{X})] \leq \mathrm{Var}[g(\mathbf{X})]$$

where equality holds if and only if $g(\mathbf{X})$ is linear.

6. The binomial distribution. Here we take as θ of Lemma 2.2 the parameter p and obtain the lower bound given in (6.1). Also, using Lemma 2.3, we find an upper bound, which, as in the Poisson case, does not have the simple form of the lower bound.

PROPOSITION 6.1. *Let X be binomial $b(n, p)$, with parameters n and p. Let $g(x)$ be a real-valued function with finite variance. Then*

$$\text{(6.1)} \quad npq\, E^2_{(n-1)}[\Delta g(X)] \leq \mathrm{Var}[g(X)] \leq npq\, E_{(n-1)}[\Delta g(X)]^2 + n^2 p \int_0^p E_{(n-1,t)}[\Delta g(X)]^2\, dt$$

where $E_{(n)}$ denotes expectations under $b(n, p)$ and $E_{(n,t)}$ denotes expectation under the binomial $b(n, t)$; $q = 1 - p$.

(This upper-bound expression is due to my Research Assistant M. Koutras).

PROOF. For the lower-bound inequality, it suffices to verify that

$$\sum_{x=0}^{n} g(x) \frac{\partial}{\partial p} \binom{n}{x} p^x q^{n-x} = n\, E_{(n-1)}[\Delta g(X)],$$

and that, as is well-known, $J_X(p) = n/pq$.

To show the upper bound, we first note that (2.3) becomes (setting again $a_k = \Delta g(k)$)

$$\text{(6.2)} \quad \mathrm{Var}[g(X)] \leq E[X \textstyle\sum_{k=0}^{X-1} a_k^2] = \sum_{k=0}^{n-1} a_k^2 \cdot \sum_{x=k+1}^{n} x \binom{n}{x} p^x q^{n-x}.$$

We evaluate the second sum here by setting $\omega = p/q$ and then differentiating both sides of the known identity (relation between the binomial tail and the incomplete beta function)

$$\sum_{x=k+1}^{n} \binom{n}{x} \omega^x = \frac{n!}{k!(n-k-1)!} (1+\omega)^n \int_0^p t^k (1-t)^{n-k-1}\, dt$$

with respect to ω. After some simplification and return to the original parameter $p = \omega(1+\omega)^{-1}$, we find

$$\sum_{x=k+1}^{n} x \binom{n}{x} p^x q^{n-x} = npq \binom{n-1}{k} p^k q^{n-k-1} + n^2 p \int_0^p \binom{n-1}{k} t^k (1-t)^{n-k-1}\, dt.$$

Therefore by (6.2) we have the upper bound in (6.1).

It should be added that, whereas the lower bound is attained if g is linear, the upper bound is attained only when g is constant (cf. the exponential case (4.2) and (5.1)).

Finally, we give a lower bound which corresponds to the situation of ν i.i.d. binomials $b(n, p)$. One easily finds that

$$\frac{npq}{\nu} \{\textstyle\sum_{k=1}^{\nu} E_k[\Delta g(X_1, \cdots, X_\nu)]\}^2 \leq \mathrm{Var}[g(X_1, \cdots, X_\nu)]$$

where Δ_k denotes the Δ operator acting only on x_k, i.e.,

$$\Delta_k g(x_1, \cdots, x_\nu) = g(x_1, \cdots, x_{k-1}, x_k + 1, x_{k+1}, \cdots, x_\nu) - g(x_1, \cdots, x_\nu),$$

and $E_k[g(X_1, \cdots, X\nu)]$ means expectation when X_ι is $b(n, p)$ for $i \neq k$ and X_k is $b(n-1, p)$.

Acknowledgement. I am indebted to Professor Herman Chernoff, who brought to my attention his inequality (1.1).

REFERENCES

CHEN, LOUIS H. Y. (1980). An inequality for the multivariate normal distribution. Technical Report, M.I.T.

CHERNOFF, HERMAN. (1980). The identification of an element of a large population in the presence of noise. *Ann. Statist.* **8** 1179–1198.

CHERNOFF, HERMAN (1981). A note on an inequality involving the normal distribution. *Ann. Probability* **9** 533–536.

MACSHANE, E. J.(1944). *Integration.* Princeton Univ. Press.

RAO, C. R. (1973). *Linear Statistical Inference and its Application*, 2nd Edition. Wiley, New York.

UNIVERSITY OF ATHENS
STATISTICAL UNIT
PANEPISTEMIOPOLIS, COUPONIA
ATHENS, GREECE