A Formula for the Expected Value of the Maximum of Three Independent Normals and a Sparse High Dimensional Case

Yiannis Kontoyiannis recently asked me if it is possible to write a formula for the expected value of the maximum of a collection of n independent, but not iid, normal variables. Since a general formula is not possible even in the iid case, unless n is small, we can only hope for a formula for small n in the independent but not iid case. For $n = 2$, a formula exists even for the general bivariate normal case, and is well known. Curiously, for $n = 3$ independent normals, a formula can be wriiten, with general means and general variances for the three normals. The derivation of this formula involves a few manipulations and tricks, and involves otherwise nothing deep. So only the formula is provided below. This is essentially the best possible closed form formula for the case of general means and general variances. Additional algebraic simplification occurs under special configurations of the means and variances. It seems extremely unlikely that an analogous formula can be written for $n = 4$ or more independent but not iid normal variables.

Let then $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), X_3 \sim N(\mu_3, \sigma_3^2)$, and suppose X_1, X_2, X_3 are independent. Let

$$\Delta_{i,j} = \mu_i - \mu_j, \rho_{i,j,k} = \frac{\sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}\sqrt{\sigma_i^2 + \sigma_k^2}}.$$

Let Φ be the standard normal CDF, ϕ the standard normal pdf, and $\overline{\Phi} = 1 - \Phi$, the tail CDF of the standard normal. Let also $\overline{\Phi_2}(u, v, \rho)$ denote the tail CDF of a bivariate normal with zero means, variances one, and correlation ρ, i.e., if (U, V) has a bivariate normal distribution with $E(U) = E(V) = 0, Var(U) = Var(V) = 1$, and correlation ρ, then $\overline{\Phi_2}(u, v, \rho) = P(U > u, V > v)$.

Define now

$$\theta(1,2,3) = \mu_1 \overline{\Phi_2}\left(\frac{\Delta_{2,1}}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \frac{\Delta_{3,1}}{\sqrt{\sigma_1^2 + \sigma_3^2}}, \rho_{1,2,3}\right) + \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \phi\left(\frac{\Delta_{2,1}}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) \overline{\Phi}\left(\frac{\Delta_{3,1} - \frac{\sigma_1^2 \Delta_{2,1}}{\sigma_1^2 + \sigma_2^2}}{\sqrt{\sigma_3^2 + \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}}}\right)$$

$$+ \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_3^2}} \phi\left(\frac{\Delta_{3,1}}{\sqrt{\sigma_1^2 + \sigma_3^2}}\right) \overline{\Phi}\left(\frac{\Delta_{2,1} - \frac{\sigma_1^2 \Delta_{3,1}}{\sigma_1^2 + \sigma_3^2}}{\sqrt{\sigma_2^2 + \frac{\sigma_1^2 \sigma_3^2}{\sigma_1^2 + \sigma_3^2}}}\right).$$

Then,

$$E(\max\{X_1, X_2, X_3\}) = \theta(1,2,3) + \theta(2,1,3) + \theta(3,2,1),$$

where $\theta(2,1,3)$ is the function obtained by interchanging the labels 1 and 2 everywhere in $\theta(1,2,3)$ and $\theta(3,2,1)$ is the function obtained by interchanging the labels 1 and 3 everywhere in $\theta(1,2,3)$.

If the means are all equal, say to some μ, then the formula reduces to an aesthetically pleasing expression, namely,

$$E(\max\{X_1, X_2, X_3\}) =$$

$$\frac{3}{4}\mu + \frac{\mu}{2\pi}[\arcsin(\rho_{1,2,3}) + \arcsin(\rho_{2,1,3}) + \arcsin(\rho_{3,2,1})] +$$

$$\frac{1}{2\sqrt{2\pi}}[\sigma_1^2(\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}} + \frac{1}{\sqrt{\sigma_1^2+\sigma_3^2}}) + \sigma_2^2(\frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}} + \frac{1}{\sqrt{\sigma_2^2+\sigma_3^2}}) + \sigma_3^2(\frac{1}{\sqrt{\sigma_2^2+\sigma_3^2}} + \frac{1}{\sqrt{\sigma_1^2+\sigma_3^2}})].$$

If, in addition, the variances are also all equal, say to some σ^2, then one retrieves from the above expression the classic formula $E(\max\{X_1, X_2, X_3\}) = \mu + \frac{3}{2\sqrt{\pi}}\sigma$.

Comment: Similar formulas should be possible for a small number of independent, but not iid, variables of other types. The discrete cases, and especially the Poisson, could be interesting to look at.

Of course, the high dimensional case is much more interesting and useful. We have no hopes of deriving exact formulas there. But perhaps we can do useful asymptotics. David Donoho and Jiashun Jin (Donoho and Jin (2004), Annals of Statistics, 962-994) gave a sparse Gaussian model, in which $100(1-\epsilon)\%$ of the observations are just noises, and another $100\epsilon\%$ are signals. They take the noises to be $N(0,1)$ and the signal to be $N(\mu,1)$, with $\mu \sim c\sqrt{2\log n}$ for some $0 < c < 1$, and ϵ to be of the order of $n^{-\beta}$ for some $\frac{1}{2} < \beta < 1$. Consider a similar, but different model, wherein exactly one in n observations is a signal (although physically unknown to the experimenter). This would correspond to $\beta = 1$ in the Donoho-Jin model; thus, here the signal is a factor of magnitude sparser than in the Donoho-Jin model. Suppose also that the signal is not really much stronger than an individual noise; i.e., μ is held fixed, rather than growing roughly at the rate of the maximum of the noisy measurements, as in the Donoho-Jin model. It thus seems that we have made the problem harder, by making the signal even sparser, and also less pronounced. But suppose that the noises are outcomes of the lab equipment, and the signal is influenced by uncontrolled factors, for example, coming from outer space. Then, the variance of the signal could be much larger than the variance of the noises in a controlled lab. Specifically, consider the model where $X_1, X_2, \cdots, X_{n-1} \overset{iid}{\sim} N(0,1)$, and $X_n \sim N(\mu, \sigma^2)$. We allow σ to depend on n (and grow with n). It is assumed that X_n is independent of the noises $X_1, X_2, \cdots, X_{n-1}$. Then, some calculation will show that $E(\max\{X_1, X_2, \cdots, X_n\})$ admits the exact formula

$$E(\max\{X_1, X_2, \cdots, X_n\}) =$$

$$\frac{(n-1)}{\sqrt{2\pi}\sigma} e^{-\frac{\mu^2}{2(1+\sigma^2)}} \int_{-\infty}^{\infty} \phi(\frac{\sqrt{1+\sigma^2}}{\sigma}(x - \frac{\mu}{1+\sigma^2}))\Phi^{n-2}(x)dx +$$

$$(n-1)(n-2)\int_{-\infty}^{\infty} \phi^2(x)\Phi(\tfrac{x-\mu}{\sigma})\Phi^{n-3}(x)dx + \tfrac{1}{\sigma}\int_{-\infty}^{\infty} x\phi(\tfrac{x-\mu}{\sigma})\Phi^{n-1}(x)dx,$$

for every fixed n and μ, σ. One may wish to know if the signal will stand out among the noises. We are currently working out the requisite asymptotics for $\max\{X_1, X_2, \cdots, X_n\}$ and its expectation.

A few values using this above exact formula may be illuminating. If $n = 50$ and $\mu = .1, \sigma = 1$, then the exact value of $E(\max\{X_1, X_2, \cdots, X_{n-1}\})$ (i.e., the largest observation among the pure noises) is 2.24119, while from the above exact formula, $E(\max\{X_1, X_2, \cdots, X_n\})$ is 2.25131. So, the signal does not stand out at all. But, if n and μ are still 50 and .1, but $\sigma = 10$, then $E(\max\{X_1, X_2, \cdots, X_n\})$ is exactly equal to 5.25534. So, there is a good chance that the signal will stand out from the noises. The theoretical asymptotics will determine when the signal will stand out in this model that is sparser than the Donoho-Jin model, and with a weaker signal, but the signal can sometimes get very large due to the higher variability. We hope to report some theoretical asymptotics soon.