Saddle-point approximations for matrix integrals: the curious case of the Wishart distribution

Alexander van Meegen

As a minimal application of the saddle-point approximation for matrix integrals we try to derive the Wishart distribution from its characteristic function using a saddle-point approximation.

A Wishart matrix $M \in \mathbb{R}^{P\times P}$ can be written as $M = \frac{1}{N}X^\top X$ where the elements of $X \in \mathbb{R}^{N\times P}$ are zero-mean Gaussian with $\langle x_{m\mu}x_{n\nu}\rangle = \delta_{nm}C_{\mu\nu}$. The corresponding distribution over symmetric positive definite (denoted by $M > 0$) matrices is [1, (3.2.1)]

$$P(M) = \frac{1}{Z}\det(M)^{\frac{N-P-1}{2}}\exp\Big(-\frac{N}{2}\operatorname{tr}C^{-1}M\Big) \tag{1}$$

where we need $N \geq P$ to ensure $C > 0$. The corresponding characteristic function is [1, (3.3.9)]

$$\Phi(\tilde{M}) = \det\Big(I_P - \frac{2i}{N}\tilde{M}C\Big)^{-\frac{N}{2}} \tag{2}$$

which is straightforward to derive from $\Phi(\tilde{M}) = \langle e^{\frac{i}{N}\operatorname{tr}\tilde{M}^\top X^\top X}\rangle_X$ by solving the Gaussian integral. Since M is symmetric, it has $P(P+1)/2$ degrees of freedom; to match the number of degrees of freedom of the characteristic function we assume that $\tilde{M}$ is symmetric as well.

Leading order

Eq. (1) and Eq. (2) are related by the inverse transformation

$$P(M) \propto \int d\tilde{M}\, e^{-i\operatorname{tr}\tilde{M}^\top M}\Phi(\tilde{M}). \tag{3}$$

Using the characteristic function we now derive Eq. (1) using a saddle-point approximation of the integral in Eq. (3) for large $N \gg 1$ and $P \gg 1$. First, we rescale $-i\tilde{M} \to \frac{N}{2}\tilde{M}$, leading to

$$P(M) \propto \int d\tilde{M}\, e^{-\frac{N}{2}S(\tilde{M})}, \qquad S(\tilde{M}) = \operatorname{logdet}(I_P + \tilde{M}C) - \operatorname{tr}\tilde{M}^\top M. \tag{4}$$

The scaling of the exponent with N suggests a saddle-point approximation. To this end, we center the integral at $S'(\tilde{M}_*) = 0$ and decompose $\tilde{M} = \tilde{M}_* + \frac{1}{\sqrt{N}}\Delta$. The saddle-point equation $S'(\tilde{M}_*) = 0$ for the above action yields

$$M = C(I_P + \tilde{M}_*C)^{-1} \qquad \Rightarrow \qquad \tilde{M}_* = M^{-1} - C^{-1} \tag{5}$$

where we used $M > 0$ and $C > 0$. Using $S(\tilde{M}_*) = -\operatorname{logdet}(M) + \operatorname{tr}(C^{-1}M) + \mathrm{const}$, changing the integration $\tilde{M} \to \Delta$, and expanding $S(\tilde{M})$ around $\tilde{M}_*$ we arrive at

$$P(M) \propto \det(M)^{\frac{N}{2}} e^{-\frac{N}{2}\operatorname{tr}(C^{-1}M)} \int d\Delta\, e^{-\frac{N}{2}\sum_{k=2}^{\infty}\frac{1}{k!N^{k/2}}\Delta^k S^{(k)}(\tilde{M}_*)} \tag{6}$$

with $\Delta^k S^{(k)}(\tilde{M}_*) \equiv \sum_{\mu_1\ldots\mu_{2k}}\Delta_{\mu_1\mu_2}\ldots\Delta_{\mu_{2k-1}\mu_{2k}}\frac{\delta^k S}{\delta\tilde{M}_{\mu_1\mu_2}\ldots\delta\tilde{M}_{\mu_{2k-1}\mu_{2k}}}\Big|_{\tilde{M}_*}$. Comparing with Eq. (1) we see that neglecting the remaining integral amounts to $\det(M)^{\frac{N-P-1}{2}} \approx \det(M)^{\frac{N}{2}}$ which is a good approximation if $N \gg P$.

Fluctuation correction

However, if P is of the same order of magnitude we need to take the fluctuations Δ into account. We do so to leading order, i.e., we expand

$$I(M) = \int d\Delta\, e^{-\frac{1}{4}\Delta^2 S''(\tilde{M}_*)} \sum_{l=0}^{\infty} \frac{1}{l!}\Big(-\frac{1}{2}\sum_{k=3}^{\infty} \frac{1}{k! N^{(k-2)/2}} \Delta^k S^{(k)}(\tilde{M}_*)\Big)^l \tag{7}$$

and neglect the terms for $l \geq 1$. The second derivative of the action is

$$S''_{\mu_1\mu_2\mu_3\mu_4}(\tilde{M}) = -C_{\mu_1\nu_1}(I_P + \tilde{M}C)^{-1}_{\nu_1\mu_3} C_{\mu_4\nu_2}(I_P + \tilde{M}C)^{-1}_{\nu_2\mu_2} \overset{\tilde{M}=\tilde{M}_*}{=} -M_{\mu_1\mu_3}M_{\mu_4\mu_2} \tag{8}$$

and thus $\Delta^2 S''(\tilde{M}_*) = -\sum_{\mu_1\ldots\mu_4} \Delta_{\mu_1\mu_2}\Delta_{\mu_3\mu_4}M_{\mu_1\mu_3}M_{\mu_4\mu_2} = -\operatorname{tr}(M\Delta M\Delta)$. To solve the integral we diagonalize $M = U\Lambda U^\top$ (recall $M > 0$) and change variables $U^\top \Delta U \to -i\Delta$, leading to

$$I(M) \propto \int d\Delta\, e^{-\frac{1}{4}\operatorname{tr}(\Lambda\Delta\Lambda\Delta)}\Big(1 + \ldots\Big) \tag{9}$$

because the orthogonal transformation has a unit Jacobian and the imaginary unit cancels with the one introduced earlier with $-i\tilde{M} \to \frac{N}{2}\tilde{M}$. We make a last change of variables $\Delta \to \Lambda^{-1/2}\Delta\Lambda^{-1/2}$ which makes the integral (to leading order) independent of M such that we only need the Jacobian. This requires care because Δ is symmetric, using [1, (1.3.5)] we arrive at

$$I(M) \propto \det(\Lambda^{-1/2})^{P+1}\Big(1 + \ldots\Big) = \det(M)^{-\frac{P+1}{2}}\Big(1 + \ldots\Big). \tag{10}$$

Inserting $I(M)$ into Eq. (6) we obtain Eq. (1) exactly.

We succeeded to derive the Wishart distribution from its characteristic function using a saddle-point approximation with fluctuation corrections. It is surprising that this is possible because we neglected higher-order terms in $I(M)$ (this is reminiscent of a similar effect happening for the chi-square distribution [2]). Roughly it seems what happens is that the Gaussian expectation of Δ^k generates terms containing M^{-k} while $S^{(k)}(\tilde{M}_*)$ generates terms containing M^k which annihilate each other at all orders such that the corrections do not depend on M.

REFERENCES

[1] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Monographs and Surveys in Pure and Applied Mathematics (Chapman & Hall/CRC, Philadelphia, PA, 1999).

[2] H. E. Daniels, Saddlepoint Approximations in Statistics, The Annals of Mathematical Statistics **25**, 631 (1954).