

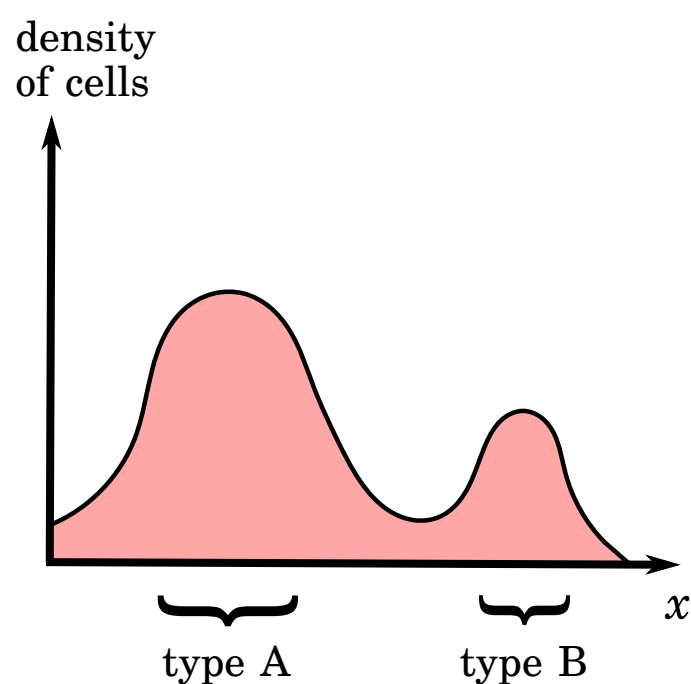


Introduction

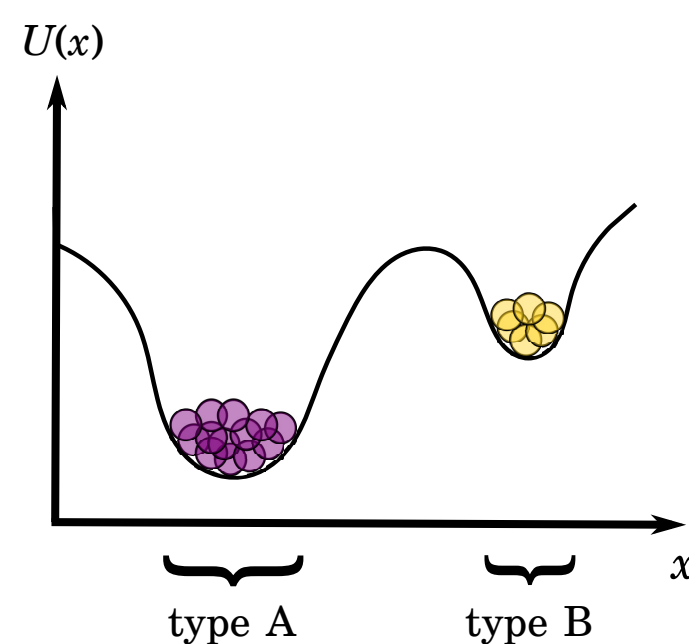
- Single-cell revolution: many data, understanding populations of cells as complex systems made of interacting individuals [1]
- Stochastic mathematics & the study of biological systems: describe cells both as individuals (random variable) and as populations (probability distribution, sample statistics)
- “Old” vs. “new” stochastic thermodynamics [2]

Epigenetic landscape & phenotypic heterogeneity [4]

probability distribution



quasipotential $U(x)$



$$x = (x_1, x_2, \dots, x_m)$$

x_i = expression level of gene i

Large deviations for mean gene expression level [3]

Suppose we measure the expression level of m genes, $X = (X_1, \dots, X_m)$, in n i.i.d. cells. Then the average gene expression profile of the sample population is given by the arithmetic mean of these samples.

Because it is itself a random variable, this sample mean is distributed according to some probability density, f_n . As the sample size n increases, the **large deviations rate function** associated with this density, $\varphi(x)$, gives us a measure of heterogeneity within the population [3]:

$$\varphi(x) = - \lim_{n \rightarrow \infty} \frac{1}{n} \ln f_n(x)$$

Equation of state for entropy function [2]

We define the **entropy** of a given sample as the rate function for the sample mean, φ , scaled by the sample size, n : $\phi(x, n) = n\varphi(x/n)$.

The function ϕ is homogeneous of order 1 in the variables (x, n) . It follows from Euler's homogeneous function theorem that the Legendre transform of $\phi(x, n)$ vanishes wherever it is finite:

$$\phi^*(\beta, \omega) = \sup_{x, n} \{ \langle \beta, x \rangle + \omega n - \phi(x, n) \} = 0,$$

where β and ω satisfy the dual relationships

$$\beta = \nabla_x \phi(x, n) \equiv \nabla \varphi(x/n), \quad \text{and} \quad \omega = \frac{\partial \phi(x, n)}{\partial n}.$$

That is, β represents the **entropic force** of gene expression, and ω represents the **information rate** associated with measuring an additional cell.

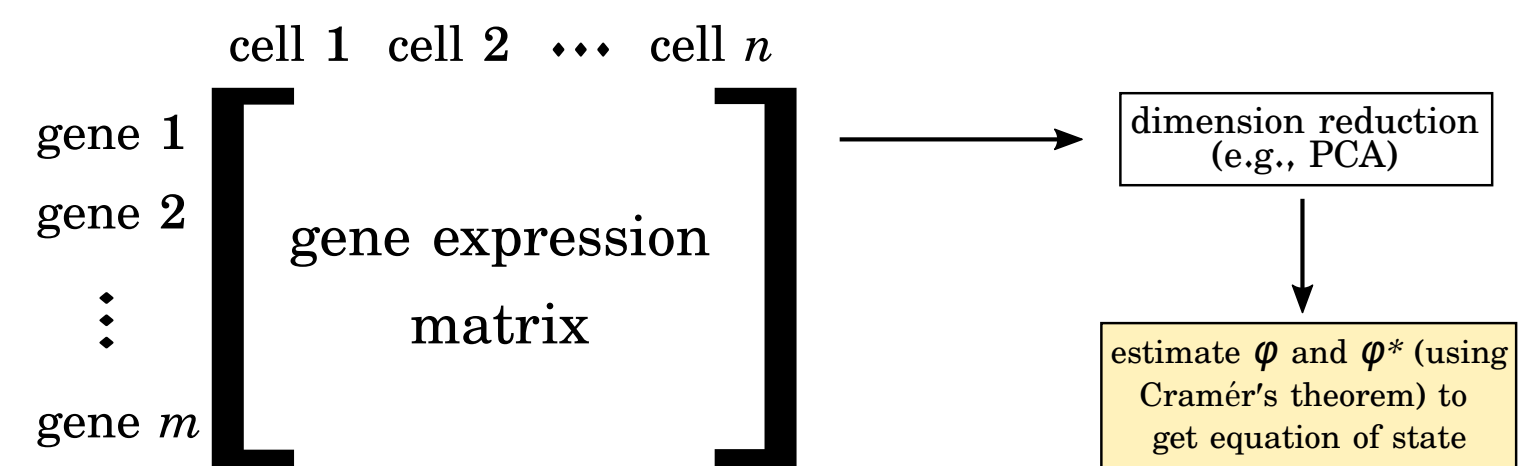
Taking the “partial” Legendre transform of ϕ in the first m variables (e.g., in x but not n) and applying $\phi^*(\beta, \omega) = 0$, we get:

$$\sup_x \{ \langle x, \beta \rangle - \phi(x, n) \} + \omega n \equiv n\varphi^*(\beta) + \omega n = 0.$$

Thus, we obtain the **equation of state** relating the conjugate variables:

$$-\varphi^*(\beta) = \omega.$$

Application to scRNA-seq data



References

- [1] S.J. Altschuler and L.F. Wu. Cellular heterogeneity: do differences make a difference? *Cell*, 141:559–563, 2010.
- [2] H. Qian. Thermodynamic behavior of statistical event counting in time: Independent and correlated measurements. *arXiv:2109.12806*, 2021.
- [3] H. Qian and Y.C. Cheng. Counting single cells and computing their heterogeneity: from phenotypic frequencies to mean value of a quantitative biomarker. *Quantitative Biology*, 8:172–176, 2020.
- [4] J.X. Zhou, M.D.S. Aliyu, E. Aurell, and S. Huang. Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface*, 9:3539–3553, 2012.