Ensemble, Estimation, Gaussian

To model "noise" or variability, we postulate an ensemble $S$ of possible observations. Consider each observation as a random draw from $S$.

For univariate quantities, we can characterize $S$ by $p(x)$, the probability that a random draw is between $x - dx$ and $x + dx$.

Typically, we are not interested in distribution $p(x)$, but rather "statistics" (Functionals) of $p(x)$, e.g.,

\[
\text{mean} = \mu = \int_S x p(x) dx = \mu_S,
\]
\[
\text{variance} = \sigma^2 = \int_S (x - \mu_S)^2 p(x) dx = \sigma^2_S.
\]
\[
\text{entropy} = -\int_S \ln p(x) dx = -\ln p(x) dx.
\]

Similarly, for multivariate $(x = (x_1, x_2, \ldots, x_N))$

\[
\text{covariance} = \Sigma = \int_S (x_j - \mu_{j,S})(x_k - \mu_{k,S}) p(x) dx.
\]

This series could consist of multivariate quantities $s(t) = (s(t_1), s(t_2), s(t_3), \ldots)$.
An estimator for a statistic $\theta$ in a procedure for taking a set of observations $X(x_1), X(x_2), \ldots, X(x_k)$ into an estimate $\hat{\theta}$.

We expect that $\hat{\theta}$ will not yield the true value $\theta = \hat{\theta}(x)$, but it should be “close.”

Two kinds of errors:

\[
\text{BIAS} = \langle \hat{\theta} - \theta \rangle_{\text{all draws of } k \text{ samples}}
\]

\[
\text{VARIANCE} = \langle (\hat{\theta} - \langle \hat{\theta} \rangle_{k \text{ draws}})^2 \rangle_{k \text{ draws}}
\]

\[
\text{VARIANCE} + (\text{BIAS})^2 = \langle (\hat{\theta} - \theta)^2 \rangle_{k \text{ draws}}
\]

The “plug-in” estimator replaces $\frac{1}{k} \sum \frac{X_{(j)}}{k}$ by $\frac{1}{k} \sum \frac{X_{(j)}}{k}$.

Plu-in estimator for means:

\[
\hat{\theta}_{\text{PI}} = k \frac{1}{k \sum_{j=1}^{k} X_{(j)}}
\]

This is unbiased, since $\langle \hat{\theta}_{\text{PI}} \rangle_{k \text{ draws}} = k \sum_{j=1}^{k} \frac{X_{(j)}}{k}$

So is the plug-in estimator for any statistic that is linear in $X$, e.g., $\sum \alpha X_r$. 

$\sum \alpha X_r$. 
The plug-in estimate for the variance $\theta_{plug}$ is:

$$\theta_{plug} = \frac{1}{K} \sum_{j=1}^{K} (X(j) - \frac{1}{K} \sum_{m=1}^{K} X(m))^2$$

$$= \frac{1}{K} \sum_{j=1}^{K} X(j)^2 - \left( \frac{1}{K} \sum_{j=1}^{K} X(j) \right)^2 - \frac{1}{K} \sum_{j=1}^{K} X(j)^2 - \left( \frac{1}{K} \sum_{m=1}^{K} X(m) \right)^2$$

If $X$ has variance 1, then $\langle \frac{1}{K} \sum_{j=1}^{K} X(j)^2 \rangle = K$ and mean 0.

$$\langle \frac{1}{K} \sum_{j=1}^{K} X(j)^2 \rangle = K$$

so $\theta_{plug} = \frac{1}{K} \cdot K - \frac{1}{K} \cdot K = 1 - \frac{1}{K}$, not 1.

So we typically use $\frac{\theta_{plug}}{K-1} = \frac{1}{K-1} \sum_{j=1}^{K} (X(j) - \frac{1}{K} \sum_{m=1}^{K} X(m))^2$ as an unbiased estimator of the variance.

The above analyses were "lucky" - bias could be calculated independent of $K$. But this is not typical. And we also don't know $J_0$ (otherwise, we wouldn't need to measure $X_{(1)}$... to estimate $\theta$).

So we need a more general setup:
Assume \( \mathcal{L} \) is part of a parametric family \( \mathcal{L}(\Theta) \).

\[ \mathcal{L}(\Theta) = \begin{cases} 1, & 1 \leq \frac{1}{\Theta} \\ 0, & \text{otherwise} \end{cases} \]

\( \Theta \) can be multivariate.

In the above case, the "plug-in" estimator for the mean is unbiased, but has unnecessarily high variance. A better estimator: choose the highest observation \( X_{\text{max}} \) and lowest \( X_{\text{min}} \); take \( \Theta = \frac{1}{10} (X_{\text{min}} + X_{\text{max}}) \).

Similarly, it's best to discard the extremes before averaging.

Nondi: plug-in estimator neither typically unbiased nor least variance.

The above set-up leads to 3 generic kinds of estimators (at least):
A. Maximum likelihood estimator

Given the observations \( x_{1}, \ldots, x_{k} \), calculate, for each \( \Theta \), the likelihood of this set of observations

\[
P(\Theta) = \prod_{m=1}^{K} p(x_{m} \mid \Theta),
\]

where each \( x_{m} \) is drawn from \( P(x \mid \Theta) \).

Find the \( \hat{\Theta}_{MLE} \) that maximizes \( P(\Theta) \).

Then \( \hat{\Theta}_{MLE} = \Theta(R(\hat{\Theta}_{MLE})) \).

B. Bayesian estimators

Assume a prior distribution for \( \Theta \), say, \( P(\Theta) \).

Given the observations \( x_{1}, \ldots, x_{k} \), the a priori distribution is modified to an a posteriori distribution

\[
P_{post}(\Theta) = \frac{P(\Theta) P(x \mid \Theta)}{\int P(\Theta) P(x \mid \Theta) d\Theta}.
\]

Con take \( \hat{\Theta}_{Bayes} = \int \Theta(R(\Theta)) P_{post}(\Theta) d\Theta \)

and

\[
\Theta_{Bayes} = \Theta(R(\hat{\Theta}_{Bayes})) \text{ where } P_{post}(\hat{\Theta}) \text{ is max}
\]

\[
\Theta_{Bayes, QAM} = \frac{\int \Theta(R(\Theta)) Q_{post}(\Theta) d\Theta}{\int Q(\Theta) Q_{post}(\Theta) d\Theta}
\]
C. "Best" estimators:
- use the procedure that minimizes
  \[ \langle (\theta - \hat{\theta}(x))^2 \rangle \]
- or, use the procedure that minimizes
  \[ \text{max} (\theta - \hat{\theta}(x))^2 \]

All of the above depend on choosing \( R(\theta) \). Usually Gaussians.

Two reasons:
- Central Limit Theorem
- Maximum Entropy property of Gaussians

For a Gaussian, the plug-in estimator is the mean, and the corrected plug-in estimator is the variance, also the MLEs.

Bayesian "best" estimators may differ; \( \hat{\theta}(x) \) could be bizarre.

Central Limit Theorem
Informally: Say \( x_i \) is drawn from \( p_2 \)
\[ x_1, \ldots, x_n \]
\[ \hat{\theta} \]

No one of these dominates. Then,
as \( k \to \infty \), \( \frac{1}{k} \sum_{j=0}^{k} x_j \) is distributed like \( n \alpha \) Gaussian.

The possibility of usage will follow some to make the contributions more precise.

Let \( y_k = x_k + x_{-k} \). The pdf \( p_k \) for \( y_k \)

\[
q_{y_k} = \int_{-\infty}^{\infty} p_{k,\omega}(y_k | x) p_{\omega}(x) dx
\]

so, with \( \hat{p}_{k|\omega} = \int e^{-i\omega y_k} p_{k,\omega} dy, dx \)

\[
\hat{q}_{\omega} = \hat{p}_{k|\omega} \hat{p}_{\omega}(\omega).
\]

Similarly for \( y_k = x_1 + \cdots + x_N \), \( \hat{q}_k(\omega) = \prod_{m=1}^{N} \hat{p}_m(\omega) \).

Let \( z_k = x_k/k, \) and \( \hat{r}_k(z) \) is the pdf of \( z_k \).

\[
\hat{r}_k(\omega) = \int e^{-i\omega z} \hat{r}_k(z) dz = \int e^{-i\omega z} q_k(kz) \cdot zdz
\]

\[
\hat{r}(\omega) = \int e^{-i\omega y_k} q_k(y_k) dy_k = \hat{q}_k(\omega) = \prod_{m=1}^{N} \hat{p}_m(\omega)
\]

\[
\log \hat{r}(\omega) = \sum_{m=1}^{k} \log \left( \hat{p}_m(\omega_m) \right)
\]
For a pair \( \nu \), \( \omega \), \( \log \left( \hat{\mu}^{\nu,\omega} \right) \) is called the 

"characteristic function" \( C_{\nu,\omega} \), obeying

\[
C_{\nu,\omega} = C_{\nu}(\omega) + C_{\omega}(\nu)
\]

Write

\[
C_{\nu}(\omega) = \sum_{s=0}^{\infty} \frac{A_s(\nu)}{s!} \omega^s \quad (\text{Interpret } \nu \text{ as } \lambda \text{ s' later})
\]

\[
\hat{\rho}(\omega) = \int e^{-iwx} \rho(x) dx, \quad \rho^{(n)}(0) = \frac{d^n}{dx^n} \rho(0) = (-i)^n \chi^n
\]

\[
\hat{\rho}(\omega) = \sum_{s=0}^{\infty} \frac{(-i)^s}{s!} \chi^s
\]

So

\[
\hat{\rho}(\omega) = 1 - i \omega \chi - \frac{\omega^2}{2} \chi^2 + \frac{\omega^3}{6} \chi^3 \ldots
\]

Let's assume that each \( \rho_m \) has mean 0. Then

\[
\hat{\rho}_m(\omega) = 1 - \frac{\omega^2}{2} \chi_m^2 + \frac{6}{\omega^3} \chi_m^3 \ldots
\]

As \( k \to \omega \), since \( \log(1 + u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \ldots \), one can approximate \( \log \hat{\rho}_m(\frac{\omega}{k}) \) as:
\[
\log \beta_m \left( \frac{\omega}{k} \right) = -\frac{\omega^2}{2k^2} \langle X_m^2 \rangle + \frac{i\omega^3}{6k^3} \langle X_m^3 \rangle + \ldots
\]

\[
- \frac{1}{2} \left( -\frac{\omega^2}{2k^2} \langle X_m^2 \rangle + \frac{i\omega^3}{6k^3} \langle X_m^3 \rangle + \ldots \right)
\]

\[
+ \frac{1}{3} (\ldots)^3 + \ldots
\]

Only one term thus is \( O \left( \frac{\omega^3}{k^3} \right) \). So

\[
\log T = \sum_{m=1}^{k} \log \beta_m \left( \frac{\omega}{k} \right) = -\frac{\omega^2}{2k^2} \langle X_m^2 \rangle + \ldots + \frac{1}{3} \left( \omega \langle X_m^3 \rangle + \ldots \right) + O \left( \frac{\omega^3}{k^3} \right)
\]

[Note: This odd \( \omega \langle X_m^3 \rangle + \ldots \) term to appear in properties of \( k \), to be sure this term dominates similarly for \( \langle X_m^3 \rangle \).]

\[
\hat{f}(\omega) \approx e^{-\frac{\omega^2}{2k^2} \langle X_m^2 \rangle} \approx e^{-\frac{\omega^2}{2k} V}
\]

where \( V \) is the average volume of the \( x_m \).

Can now estimate \( r(z) = \frac{1}{2\pi i} \int_{C} e^{iz\hat{f}(\omega)} d\omega \)

\[
\approx \frac{1}{\sqrt{2\pi} k} e^{-\frac{1}{2} z^2/(V/k)}
\]

(a Gaussian of \( \frac{V}{k} \), we'll do these integrals later) volume \( \frac{V}{k} \).

[Note fine point: \( \gamma \) in \(*\) more: second moment well approximated, higher moments not well approximated.***]
Interpretation of the A's - $A_{k}(p)=A_{k}(p) + A_{k}(q)$

\[ \log \gamma (\mu \omega) = \sum_{k} \gamma_{k}(\omega) \gamma_{k}(\mu) \]

\[ = - \frac{1}{\omega} \left( \sum_{k} \frac{1}{\omega} \frac{\omega^{2}}{k^{2}} \right) + \frac{1}{\omega^{3}} \left( \sum_{k} \frac{1}{\omega} \frac{\omega^{2}}{k^{2}} \right) \cdots \]

\[ - \frac{1}{\omega} \left( \sum_{k} \frac{1}{\omega} \frac{\omega^{2}}{k^{2}} \right) \cdots \]

\[ + \frac{1}{\omega} \left( \sum_{k} \frac{1}{\omega} \frac{\omega^{2}}{k^{2}} \right) \cdots \]

\[ + O(\omega) \]

\[ A_{1} = - i \omega \chi \Rightarrow A_{1} = \omega \chi \]

\[ \frac{A_{2}}{\omega} = - \frac{1}{\omega} \omega^{2} \chi^{2} + \frac{1}{2} \omega \chi^{2} \Rightarrow A_{2} = \omega \chi \chi^{2} - \chi \chi^{2} \]

\[ \frac{A_{3}}{\omega} = \frac{1}{\omega} \omega \chi^{2} - \frac{1}{2} \omega \chi \chi^{2} + \frac{1}{3} \omega^{3} \chi^{3} \]

\[ A_{3} = \omega \chi^{2} - 3 \omega \chi \chi^{2} + 2 \omega \chi^{3} \]

$A_{1}$ = main, $A_{2}$ = variance, $A_{3}$ = skewness, $A_{4}$ = kurtosis
Some Gaussian integrals.

\[ I = \int \cdots \int e^{-\frac{1}{2} (\mathbf{x} - \mathbf{\alpha})^T M (\mathbf{x} - \mathbf{\alpha})} \, d\mathbf{x} \]

where \( M \) is a \( K \times K \) matrix, symmetric.

\[ (\mathbf{x} - \mathbf{\alpha})^T M (\mathbf{x} - \mathbf{\alpha}) = \sum_{i,j=1}^{K} (x_i - \alpha_i)(x_j - \alpha_j)M_{ij} \]

Say we compute \( M = R D R^{-1} \).

Say \( R \) real unitary, \( D \) diagonal with positive eigenvalues.

\[ (\mathbf{x} - \mathbf{\alpha})^T M (\mathbf{x} - \mathbf{\alpha}) = \mathbf{y}^T D \mathbf{y}, \quad \mathbf{y} = (\mathbf{x} - \mathbf{\alpha}) R \]

[If \( D \) had an eigenvalue \( \leq 0 \), the integral would not \( \to 0 \) \( \Rightarrow \) \( \mathbf{y} \) in a common eigenvector.]

Say \( D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_K \end{pmatrix} \).

\[ \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \\ \vdots \\ y_K \end{pmatrix}, \quad y_i = \sqrt{\lambda_i} \]

\[ \mathbf{z} = \mathbf{y} \sqrt{D} = (\mathbf{x} - \mathbf{\alpha}) R \sqrt{D} \]

\[ \mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_k \\ \vdots \\ z_K \end{pmatrix}, \quad z_i = \sqrt{\lambda_i} \]

\[ (\mathbf{x} - \mathbf{\alpha})^T M (\mathbf{x} - \mathbf{\alpha}) = \mathbf{z}^T \mathbf{z}, \quad \mathbf{z} = \mathbf{y} \sqrt{D} \]

\[ \mathbf{z}^* = \frac{1}{\sqrt{\det M}} \mathbf{z} \quad (\det R = 1). \]
\[ I = \int \ldots \int e^{-\frac{x+1}{2}} e^{\left(\frac{a^T A^{-1} b}{2}\right)} \frac{dx^2}{\sqrt{\det M}} \]

\[ = \frac{e^{a \cdot b}}{\sqrt{\det M}} \int \ldots \int e^{-\frac{\sum_{i=1}^{n} z_i^2}{2}} e^{\sum_{m=1}^{k} c_m z_m} \frac{dz_1 \ldots dz_k}{2} \]

\[ = \frac{e^{a \cdot b}}{\sqrt{\det M}} \prod_{m=1}^{k} \int e^{-\frac{z_m^2}{2} + c_m z_m} \frac{dz_m}{2} \]

\[ = \frac{e^{a \cdot b}}{\sqrt{\det M}} \prod_{m=1}^{k} e^{c_m z_m/2} \int e^{-z_m^2/2} e^{-z_m^2/2} \frac{dz_m}{2} \]

\[ = \frac{e^{a \cdot b}}{\sqrt{\det M}} \prod_{m=1}^{k} e^{c_m z_m/2} \left[ \int e^{-u^2} du = \sqrt{\pi} \right] \]

\[ \sum c_m^2 = \sum c_m z_m = c \cdot \sum (b \cdot f_m^T) (A^{-1} b \cdot f_m^T) = bM^{-1} b^T \]

\[ I = \frac{e^{a \cdot b}}{\sqrt{\det M}} \prod_{m=1}^{k} e^{c_m z_m/2} \cdot bM^{-1} b^T/2 \]
\[ M = V^{-1}, \quad \epsilon = 0, \quad l = 0. \]

So,
\[ \int \int e^{-\frac{x^2}{2V^2x^2/2}} \, dx = \frac{(2\pi)^{k/2}}{\sqrt{\det V}}, \quad \int \int e^{-\frac{x^2}{2V^2x^2/2}} \, dx = 1. \]

Similarly, \( V_{kk} = V_{mm} \)

\[ \int \int x_kx_m e^{-\frac{x^2}{2V^2x^2/2}} \, dx = V_{km} \quad (\dagger) \]

Thus establish \( \frac{1}{\sqrt{(2\pi)^k}} \frac{1}{\sqrt{\det V}} e^{-\frac{x^2}{2V^2x^2/2}} \) as the p.d.f. of a Gaussian with \( V_{kk} \) given by \( \text{cov} \times \text{cov}^{-1} \times V = \frac{1}{2} V_{km} S. \)

* Similar argument: \[ x^T x = (\frac{1}{\sqrt{\det B}} R^T) \frac{1}{\sqrt{\det B}} x \cdot \frac{1}{\sqrt{\det B}} R^T \]

\[ \int \int \frac{1}{\sqrt{(2\pi)^k}} \frac{1}{\sqrt{\det V}} e^{-\frac{x^2}{2V^2x^2/2}} \, dx = \left( \frac{1}{\sqrt{(2\pi)^k}} \frac{1}{\sqrt{\det V}} \int \int e^{-\frac{x^2}{2}} \, dx \right) \sqrt{\det V} \frac{1}{\sqrt{\det B}} R^T \]

\[ \begin{pmatrix} \frac{1}{\sqrt{(2\pi)^k}} \frac{1}{\sqrt{\det V}} \int \int e^{-\frac{x^2}{2}} \, dx \end{pmatrix} \sqrt{\det V} \frac{1}{\sqrt{\det B}} R^T = \frac{1}{\sqrt{(2\pi)^k}} \frac{1}{\sqrt{\det V}} \int \int e^{-\frac{x^2}{2}} \, dx \cdot \frac{1}{\sqrt{\det B}} R^T = M^{-1} = V \]
If \( p(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \frac{1}{\det V} e^{-\frac{x^T V^{-1} x}{2}} \),

\[
\hat{p}(\omega) = \int \int p(x) e^{-i \omega^T x} dx = e^{-i \omega^T \mu} \quad \left[ \text{see page 4} \right]
\]

\[ e^{-i \omega^T \mu} \left[ \begin{array}{c} \mathbf{E} = V^{-1} \\ \mu = i \omega \\ bT = -i \omega T \end{array} \right] \]

**Maximum Entropy Distributions**

It would be nice to specify \( P(\omega) \) with only a few parameters—n in particular, and like not to have to specify all moments, or, all PDF's.

Entropy of a discrete distribution \( p(x) \)

\[ p(x) = \sum_{r=1}^{n} p_r \delta(x-x_r) \]

given by \( H(p) = -\sum p_r \log p_r \)

If \( \log = \log_2 \), \( H(p) \) is the yes–no questions, on average, needed to determine \( r \), if each obs. occurs with frequency \( p_r \).
For a continuous distribution:

\[ H(p) = \lim_{\Delta x \to 0} \sum_{x_i} (p(x_i) \Delta x) \log p(x_i; \Delta x) \]

\[ = \lim_{\Delta x \to 0} \sum_{x_i} p(x_i) \Delta x \left( \log p(x_i) + \log \Delta x \right) \]

\[ = \lim_{\Delta x \to 0} \left[ \sum_{x_i} p(x_i) \log p(x_i) \Delta x + \log \Delta x \sum_{x_i} p(x_i) \Delta x \right] \]

\[ = \int p(x) \log p(x) dx + \lim_{\Delta x \to 0} \left( \log \Delta x \right) \]

"Differential entropy" of \( p \)

"Nuisance" is irrelevant for comparing entropies.

Plan: Specify only a few moments of \( p \), by \( \zeta \).

Then, find \( \beta(\zeta) \) as maximum differential entropy ensemble with moments \( \zeta \).
How to find a maximum distribution, given some constraints?

Discrete case:

Entropy: \( H = - \sum p_r \log(p_r) \)

Typical constraint: \( \text{Mean} = \mu_r \Rightarrow \sum p_r x_r = \mu \)

Variance: \( \sigma^2 \Rightarrow \sum p_r (x_r - \mu)^2 = \sigma^2 \)

The constraints are linear in \( p_r \).

\( k \) th constraint: \( \sum p_r c_{kr} = A_k \)

[We even have a 0th constraint: \( \sum p_r = 1 \), i.e., \( C_{0k} = 1 \)]

Method of Lagrange Multipliers:

Set \( f = f(u_1, u_2, \ldots, u_n) \) to be maximized, subject to constraints \( g_i(u_1, \ldots, u_n) = a_k \).

Let \( F = f(u_1, \ldots, u_n) + \sum_{k=1}^{K} \lambda_k g_k(u_1, \ldots, u_n) \).

The constraint extremum of \( F \) occurs where \( \frac{\partial F}{\partial u_j} = 0 \).

This typically leads to equations \( \lambda_k = \lambda_l \) for \( k \neq l \), from which the values of the \( \lambda \)'s must be determined.
For entropy, the constrained maximum, if it exists, must be unique.

Say $p + q$ are both local maxima. They satisfy the constraints with $H(p) = H(q)$ but $p 
eq q$. 

\[ H((1 - m)p + mq) > H(p) \text{ or } H(q) \]

[Mixing property of entropy.]

\[ H \]

\[ p \]

\[ q \]

\[ \text{[Insights must be]} \]

\[ \text{[Convex]} \]

Apply Lemma: Multiply to 
\[ -\frac{\partial}{\partial p_j} \log p_j - \frac{\partial}{\partial p_j} \sum_k p_k C_{kj} \]

\[ F = -\frac{1}{m} \log p_j + \frac{\xi}{\alpha} \sum_k p_k C_{kj} \]

\[ \frac{\delta F}{\delta p_j} = -1 + \log p_j + \frac{\xi}{\alpha} \sum_k C_{kj} \]

\[ p_j = B e^{\frac{\xi}{\alpha} \sum_k C_{kj}} \quad \text{with } C_{00} = 1 \]

So, if we constrain variances + many, the $C_{kj}$ will be

\[ C_{00} = \frac{x_r}{\alpha} \]

\[ C_{22} = \frac{x_r^2}{\alpha^2} \]
Continuous univariate case:
\[ \mathcal{E} = - \int p(x) \log p(x) \, dx + \sum_{k=0}^{K} \int p(x) \ C_k(x) \, dx \]

\( C_k \rightarrow C_k(x) \) \quad \text{with} \quad \frac{\partial \mathcal{E}}{\partial \left[ C_k(x) \right]} \quad \text{[formal, solution]}

\[ \frac{\partial \mathcal{E}}{\partial [p(x)]} = -1 - \log p(x) + \sum_{k=0}^{K} \int p(x) \ C_k(x) \, dx \]

and we have:
\[ p(x) = \mathcal{E} \]
\[ \mathcal{E} = \sum_{k=0}^{K} \int p(x) \ C_k(x) \, dx \]

\[ \text{Prior:} \quad C_0(x) = 1 \]
\[ \text{Verify:} \quad C_1(x) = x \]

Case \( p(x) \) is a Gaussian, we need to choose \( C_1 = C_2 \).

Easiest to first shift the problem so that \( \mu = 0 \), and then
use (4) \( \mathcal{E} \) to get \( \mathcal{E} = \frac{1}{2} V \).

Multivariate case: once control for each variable's mean
\[ \mu \quad \text{covariance} \]

Some strategy:

Conclusion: The multivariate Gaussian is the natural distribution for a constrained mean + covariance.