Entropy in Asset Pricing Research and Beyond:
A Large Deviations Approach
With Potential to Classify Model Complexities

Michael Stutzer
Professor of Finance
University of Colorado
Boulder, CO

I applied Entropic Transformation of Probabilities to 3 areas of Securities Price Research

1. Derivative Securities Pricing (e.g. options traded on the CBOE) [Stutzer, *J. of Finance*]
    - Concept: Transform the securities' return probabilities into a *risk-neutral measure* that satisfies the no-arbitrage constraints on prices.
        - Use the resulting measure to price the derivative securities.
2. Estimation and Testing of Asset Pricing Models [Kitamura&Stutzer, *Econometrica*, *J. of Econometrics*]
    - Concept: For a given possible values of the parameters, transform the data generating mechanism probabilities into a measure that satisfies the moment constraints implied by the model.
    - Search over parameter values to find the transformed measure entropically closest to the data generating distribution.
        - Adopt those parameter values as parameter estimates, and test the model based on the aforementioned entropic closeness.
3. Optimal Portfolio Choice for Long-Term Investment [Stutzer, *J. of Econometrics*]
    - Concept: Find "risk-controlled" portfolios that make shortfall probabilities go to zero as fast as possible.
    - This also uses an entropically-transformed return distribution, but the reason for this is rooted in a frequentist interpretation of entropy that arises in the *Statistical Theory of Large Deviations*.

Ol' Reliable: Entropic Solution of Linear Inverse Problem

- Define *entropy D of Q relative to P (a.k.a. Kullback-Leibler)*:

$$D(Q,P) = E^Q[\log \frac{dQ}{dP}]$$

$$\overset{discrete}{=} \sum_j Q_j \log \frac{Q_j}{P_j}$$

- Find measure $Q^* \triangleq \arg\min_{Q_1,\ldots,Q_N} \sum_j Q_j \log \frac{Q_j}{P_j} \ s.t. \sum_j Q_j = 1, \ \ and$

$$E^Q[X_1] \equiv \sum_j X_{1j} Q_j = c_1$$

$$E^Q[X_2] \equiv \sum_j X_{2j} Q_j = c_2$$

$$\vdots$$

$$E^Q[X_r] \equiv \sum_j X_{rj} Q_j = c_r$$

In my 3 Applications:

1. Derivative Security Price P with one underlying asset:

$$X_1 \triangleq X_1(T) / X_1(0) \textit{ is the underlying security's cumulative return}$$

$$E^Q[X_1] = c_1 \triangleq e^{rT} \textit{ where } r \textit{ is the interest rate } (\text{"}Risk-Neutral\text{"}\ Constraint)$$

$$Q^* \textit{ is the estimated risk-neutral pricing measure}$$

$$P = E^{Q^*}[payoff[X_1(T)]] / e^{-rT} \textit{ is the predicted derivative price}$$

2. Estimation of parameters θ and Testing of Asset Pricing Models

$$E^P[X_i; \theta_{True}] = c_i \textit{ is the model's } i^{th} \textit{ moment restriction } (\textit{fails when } \theta \neq \theta_{True})$$

$$\hat{\theta} \triangleq \arg\min_\theta \min_Q D(Q, P) \ s.t.\ E^Q(X_i; \theta) = c_i,\ \ i = 1, \dots, r$$

$$\textit{is the "exponential tilting" estimator}$$

The 3rd Application

3. Portfolio Choice θ^* that Minimizes Asymptotic Shortfall Probabilities

$$E^Q[X_1;\theta] \triangleq \textit{Expected Log Gross Return of Portfolio With Weights } \theta$$

$$c_1 \triangleq \textit{shortfall threshold return} < E^P[X_1;\theta]$$

$$\theta^* = \arg\max_\theta \min_Q D(Q,P) \; s.t. \; E^Q[X_1;\theta] = c_1$$

Q: But *why* is θ^* a vector of shortfall minimizing portfolio weights?
A: *Statistical Theory of Large Deviations* explains why.

The Frequentist Interpretation of $\text{Min}_Q\, D(Q,P) = D(Q^*,P)$

- In empirical problems with *T observations on the vector X* used to estimate *E[X]*, the LLN ensures that the sample mean vector under *Q* will converge to the vector of *Q*-expected values specified by the constraints. <u>But under the *P*-measure</u>:

$$Prob\left[\sum_{t=1}^{T} \vec{X}_t / T \le \vec{c}\right] \approx \frac{k}{\sqrt{T}}\, e^{-I(\vec{c})T} \quad \textit{for large } T$$

$$I(\vec{c}) \triangleq \max_{w_1,\ldots,w_r} \sum_{i=1}^{r} w_i c_i - \log E^P[e^{\sum_{i=1}^{r} w_i X_i}] \quad (\textit{Cramer's Theorem})$$

$$= D(Q^*, P) \quad (\textit{Kullback's Lemma})$$

$$Q_j^* = \frac{e^{\sum_{i=1}^{r} w_i^* X_{ij}} P_j}{\sum_{j=1}^{N} e^{\sum_{i=1}^{r} w_i^* X_{ij}} P_j}$$ are conditional probabilities *given* $\bar{X} \approx \vec{c}$

Large Deviation Rate Functions $I(\vec{c})$

- For IID Processes (Cramer’s Theorem): Legendre Transform of Cum.Gen.Function

$$I(\vec{c}) \triangleq \max_{w_1,\ldots,w_r} \sum_{i=1}^{r} w_i c_i - \log E[e^{\sum_{i=1}^{r} w_i X_i}] = \sum_{i=1}^{r} w_i c_i - CGF(w_1,\ldots,w_r) = D(Q^*, P)$$

- For Non-IID, Weakly Dependent Processes (Gartner-Ellis Theorem)

$$I(\vec{c}) \triangleq \max_{w_1,\ldots,w_r} \sum_{i=1}^{r} w_i c_i - \frac{1}{T}\log E[e^{\sum_{i=1}^{r} w_i \sum_{t=1}^{T} X_{it}}] = \sum_{i=1}^{r} w_i c_i - Scaled\ CGF(w_1,\ldots,w_r)$$

- So *define* this Legendre Transform of the *Scaled* Cumulant Generating Function this as a *generalized entropy*, which reduces to K-L entropy when the process is IID.
- *Using this Large Deviations interpretation, constrained entropy optimization can be rationalized without resort to axiomatic, information-theoretic viewpoints*.

This Frequentist Interpretation Rationalizes the Other Applications, too.

- Kitamura and Stutzer exposit the Large Deviations frequentist interpretation of the exponential tilting estimator using both of these results.

- Kitamura also used Large Deviations to provide a frequentist foundation for *empirical likelihood estimators, resulting from replacing* $D(Q,P)$ *with* $D(P,Q)$

- *More generally*, this reasoning reflects the notion of *Bahadur Efficiency*.

Relevance of this for Complexity Conference Questions #1 and #2

- Both equilibrium and non-equilibrium Stat Mech models of complex interaction have modern frequentist foundations based on properties of their respective large deviations rate functions
  - The *Least Complex* Stat Mech models have strictly convex rate functions $I(\vec{c})$ with a unique equilibrium state associated with $\min_c I(\vec{c}) = 0$.
  - *More Complex* Stat Mech models have non-convex rate functions $I(\vec{c})$, possibly with multiple equilibrium states associated with different critical points of $I(\vec{c})$. *Such phenomena result in more complex behavior, e.g. phase transitions*.
  - This provides a unified classification of the properties of Stat Mech models, as developed in Touchette (*Physics Reports*, 2009)

A Proposal:

- For constrained entropy models in the social sciences that admit this frequentist interpretation, classify their complexity by analogous characterization of their respective rate functions.

- That is, derive *their respective rate functions* $I(\vec{c})$ *, and classify in accord with several properties:*
  1. *Number of parameters required for computation of* $I(\vec{c})$
  2. *The number of critical points of* $I(\vec{c})$
  3. *The nature of associated non-convexities, e.g. are their linear segments in* $I(\vec{c})$ *, or non-differentiable points in the domain?*
  4. *How the above properties vary across the free parameter space.*

- Try using this information to construct an index of complexity. Other indices do exist, based on algorithmic information theory (e.g. Gell-Mann and Lloyd, 1996).
  - But these are difficult to calculate, and I doubt this has led to many practical insights.