lecture04

December 18, 2024

[2]:
```
b = [1/3, 1/2, 1/6];
o = [5, 3, 10];
```

[6]:
```
p(x) = x -> b(x) * o(x);
```

[7]:
```
@show p(1);
```

```
p(1) = var"#3#4"()
```

Doubling Rate of a horse race

$W(b,p) = E(logS(X)) = \sum_{k=1}^{m} p_k log b_k o_k$

The given formula

$J(b) = \sum_i p_i \log(b_i \cdot o_i) + \lambda \sum_i b_i$

represents an **objective function** for optimization in a probabilistic betting or resource allocation problem, where the Lagrange multiplier λ enforces a constraint on the allocation of resources b_i. Here's how it can be interpreted and optimized:

0.0.1 Components of the Formula

1. **Objective Function (Logarithmic Growth):** $\sum_i p_i \log(b_i \cdot o_i)$
   - p_i: Probability of outcome i,
   - b_i: Fraction of the total resource (e.g., wealth) allocated to outcome i,
   - o_i: Odds associated with outcome i,
   - This term maximizes the **expected logarithmic growth** of wealth.
2. **Constraint Term (Resource Allocation):** $\lambda \sum_i b_i$,
   - λ: Lagrange multiplier, enforces the constraint on the allocation $\sum b_i = 1$, ensuring that the total betting fraction equals the available resource (e.g., all wealth is distributed across outcomes).

0.0.2 Optimization with Lagrange Multipliers

Step 1: Define the Full Objective The **Lagrangian** becomes: $\mathcal{L}(b,\lambda) = \sum_i p_i \log(b_i \cdot o_i) + \lambda \left(1 - \sum_i b_i\right)$, where: - The term $((1 - \sum b_i))$ ensures the constraint $\sum b_i = 1$ is enforced.

Step 2: Compute the Gradient To find the optimal b_i and λ, set the partial derivatives of $\mathcal{L}$ to zero:

1. **Derivative with respect to** b_i**:** $\frac{\partial \mathcal{L}}{\partial b_i} = \frac{p_i}{b_i} + \lambda = 0$.
2. **Derivative with respect to** λ**:** $\frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_i b_i = 0$.

Step 3: Solve for b_i From the first condition: $b_i = -\frac{p_i}{\lambda}$.

Substitute this into the second condition: $\sum_i b_i = \sum_i -\frac{p_i}{\lambda} = 1$.

Solve for λ: $\lambda = -\sum_i p_i = -1 \quad (\text{since} \sum p_i = 1)$.

Thus: $b_i = p_i$.

0.0.3 Result

The optimal betting fractions b_i are proportional to the probabilities p_i, aligning the betting strategy with the actual probabilities of outcomes. This result maximizes the expected logarithmic growth while satisfying the constraint $\sum b_i = 1$.

Example:

- Consider m = 2 horses
- With probability of winning P_1, P_2.
- Even odds (2-for-1 on both horses - $o_i = 2$)
- Optimal (proportional) betting

$b_1 = p_1$, $b_2 = p_2$

- The optimal doubling rate is
- $W^*(p) = \sum_{i=1}^{m} p_i log o_i - H(p) = 1 - H(p)$

The **doubling rate** measures the expected logarithmic growth of wealth when betting optimally on the horses. Given the setup:

0.0.4 Parameters:

1. **Number of horses:** $m = 2$,
2. **Probabilities of winning:** p_1 and p_2 (with $p_1 + p_2 = 1$),
3. **Odds:** $o_1 = o_2 = 2$ (even odds),
4. **Optimal betting fractions:** $b_1 = p_1$, $b_2 = p_2$.

0.0.5 Formula for the Optimal Doubling Rate:

The **optimal doubling rate** $W^*(p)$ is given by: $W^*(p) = \sum_{i=1}^{m} p_i \log(o_i) - H(p)$, where: - $H(p)$ is the Shannon entropy of the probability distribution: $H(p) = -\sum_{i=1}^{m} p_i \log(p_i)$, - The term

$\sum_{i=1}^{m} p_i \log(o_i)$ captures the expected logarithmic return based on the odds.

0.0.6 Step-by-Step Calculation:

1. Logarithmic Return ($\sum p_i \log(o_i)$): $\sum_{i=1}^{m} p_i \log(o_i) = p_1 \log(2) + p_2 \log(2)$. Since $\log(2) = 1$: $\sum_{i=1}^{m} p_i \log(o_i) = p_1 \cdot 1 + p_2 \cdot 1 = p_1 + p_2 = 1$.

2. Shannon Entropy ($H(p)$): $H(p) = -\left(p_1 \log(p_1) + p_2 \log(p_2)\right)$.

3. Combine Terms: $W^*(p) = \sum_{i=1}^{m} p_i \log(o_i) - H(p)$.

Substitute the results:

$W^*(p) = 1 - \left(-p_1 \log(p_1) - p_2 \log(p_2)\right)$, $W^*(p) = 1 + p_1 \log(p_1) + p_2 \log(p_2)$.

0.0.7 Final Doubling Rate:

The optimal doubling rate is: $W^*(p) = 1 - H(p)$, where $H(p)$ is the Shannon entropy: $H(p) = -p_1 \log(p_1) - p_2 \log(p_2)$.

0.0.8 Example Calculation:

Suppose: - $p_1 = 0.6$, $p_2 = 0.4$, - Odds $o_1 = o_2 = 2$.

1. Compute $H(p)$: $H(p) = -(0.6 \log(0.6) + 0.4 \log(0.4))$. Using log base 2:

- $\log(0.6) \approx -0.737$,
- $\log(0.4) \approx -1.322$, $H(p) = -(0.6 \cdot -0.737 + 0.4 \cdot -1.322)$,

$H(p) = 0.6 \cdot 0.737 + 0.4 \cdot 1.322 = 0.442 + 0.529 = 0.971\,\text{bits}$.

2. Compute $W^*(p)$: $W^*(p) = 1 - H(p) = 1 - 0.971 = 0.029\,\text{bits}$.

0.0.9 Interpretation:

- The optimal doubling rate is $W^*(p) = 0.029\,\text{bits}$ for this scenario.
- This means your wealth grows by a factor of $2^{0.029} \approx 1.02$ per round on average, under optimal betting.

$D(P||r) - D(P||b) = W$

r is bookee's understanding
b is gambler's understanding

The equation:

$W^*(p) + H(p) = \log m$

relates the **optimal doubling rate** $W^*(p)$, the **Shannon entropy** $H(p)$, and the total number of possible outcomes m in a probabilistic betting scenario. Here's what it represents:

0.0.10 Key Components

1. m**:**
   - The total number of possible outcomes in the system. For example, if you are betting on a horse race with $m = 2$ horses, there are two possible outcomes.
2. **Shannon Entropy** $H(p)$**:**
   - Measures the **uncertainty** or randomness of the probability distribution $p = \{p_1, p_2, \dots, p_m\}$.
   - Defined as: $H(p) = -\sum_{i=1}^{m} p_i \log p_i$.
3. **Optimal Doubling Rate** $W^*(p)$**:**
   - Represents the maximum expected logarithmic growth rate of wealth when betting optimally. For a given probability distribution p, it is: $W^*(p) = \sum_{i=1}^{m} p_i \log o_i - H(p)$, where o_i are the odds. Under **fair odds** ($o_i = 1/p_i$), this simplifies to: $W^*(p) = \log m - H(p)$.

0.0.11 Interpretation of the Equation

From the fair odds scenario: - The total information content (logarithm of the number of outcomes, $\log m$) is split into two parts: 1. $H(p)$: The **entropy** or randomness inherent in the probability distribution p, 2. $W^*(p)$: The remaining information that can be exploited for wealth growth through optimal betting.

Thus: $W^*(p) + H(p) = \log m$ means that the **maximum doubling rate** plus the entropy equals the total "information potential" of the system.

0.0.12 Special Cases

1. Uniform Distribution: If $p_i = 1/m$ (all outcomes are equally likely): - Entropy is maximized: $H(p) = \log m$. - $W^*(p) = 0$, since there is no exploitable edge under fair odds.

Thus: $W^*(p) + H(p) = \log m \implies 0 + \log m = \log m$.

2. Deterministic Distribution: If $p_i = 1$ for one outcome and $p_j = 0$ for all others: - Entropy is minimized: $H(p) = 0$. - Doubling rate is maximized: $W^*(p) = \log m$.

Thus: $W^*(p) + H(p) = \log m \implies \log m + 0 = \log m$.

0.0.13 Conclusion

The equation $W^*(p) + H(p) = \log m$ highlights the balance between the **entropy** of the system (uncertainty) and the **optimal growth rate** (exploitable edge). It encapsulates the interplay between randomness and strategy in probabilistic systems, especially in the context of optimal betting under fair odds.

The **entropy rate** is a measure of the **average uncertainty per symbol** in a **stochastic process** or sequence of random variables. It generalizes the concept of entropy to sequences or time series, capturing the long-term average information content per observation.

0.0.14 Definition

Given a stochastic process $\{X_1, X_2, \dots\}$ consisting of random variables, the **entropy rate** $H(X)$ is defined as:

$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$,

where: - $H(X_1, X_2, \dots, X_n)$ is the joint entropy of the first n variables in the sequence, - The limit ensures we capture the long-term average.

0.0.15 Key Scenarios

1. **Independent and Identically Distributed (IID) Process:**
   - If $\{X_t\}$ is an IID process (each X_t is independent with the same marginal distribution), the entropy rate is simply the entropy of one variable: $H(X) = H(X_1)$.
2. **Markov Process:**
   - For a first-order Markov process, where $P(X_n|X_{n-1}, \dots, X_1) = P(X_n|X_{n-1})$: $H(X) = H(X_2|X_1)$.
   - The entropy rate depends only on the conditional entropy of the current state given the previous one.
3. **Stationary Processes:**
   - For stationary processes (where the probability distributions do not change over time), the entropy rate is well-defined and reflects the average uncertainty per symbol.

0.0.16 Alternative Formulation

The entropy rate can also be expressed in terms of **conditional entropy**: $H(X) = \lim_{n\to\infty} H(X_n|X_{n-1}, X_{n-2}, \dots, X_1)$, which measures the uncertainty of the current symbol given all past symbols.

0.0.17 Properties

1. **Units:** The entropy rate is measured in bits (if using base 2 logarithms) or nats (if using natural logarithms).
2. **Bounds:** $0 \leq H(X) \leq \log|\mathcal{X}|$, where $\mathcal{X}$ is the alphabet size of the process.
3. **IID Case:** $H(X)$ is maximal when the symbols are IID and uniformly distributed.

0.0.18 Example Calculations

1. IID Process: Let $\{X_t\}$ be IID with $P(X = 0) = 0.5$ and $P(X = 1) = 0.5$. The entropy rate is: $H(X) = H(X_1) = -[0.5 \log 0.5 + 0.5 \log 0.5] = 1\,\text{bit}$.

2. First-Order Markov Chain: Consider a binary Markov chain with states $\{0, 1\}$ and transition probabilities: $P(X_n = 1|X_{n-1} = 0) = 0.8, \quad P(X_n = 0|X_{n-1} = 0) = 0.2$. The entropy rate is: $H(X) = H(X_2|X_1) = \sum_{x,x'} P(X_1 = x, X_2 = x') \log P(X_2 = x'|X_1 = x)$.

3. Stationary Source: For a stationary source generating symbols $\{A, B, C\}$ with probabilities $P(A) = 0.5, P(B) = 0.3, P(C) = 0.2$, and no memory (independence): $H(X) = H(X_1) = -(0.5 \log 0.5 + 0.3 \log 0.3 + 0.2 \log 0.2)$.

0.0.19 Applications

1. **Data Compression:**
   - The entropy rate sets the theoretical limit for the average number of bits required to encode the stochastic process.
2. **Information Theory:**
   - The entropy rate quantifies the information content of a source emitting a sequence of random variables.
3. **Statistical Learning:**
   - Understanding entropy rates helps in modeling time-series data and estimating the complexity of dynamical systems.
4. **Cryptography:**
   - Processes with higher entropy rates are more secure due to increased randomness.

0.0.20 Summary

The **entropy rate** captures the long-term average uncertainty per symbol in a stochastic process. For IID processes, it equals the entropy of a single random variable, while for dependent or stationary processes, it reflects the influence of past observations. The entropy rate is a key concept in compression, signal processing, and time-series analysis.

Example (Red and Black)

0.0.21 Given Formula:

The value gain is tied to:

$S^*_{52} = \frac{2^{52}}{\binom{52}{26}}$,

where: - 2^{52}: Total number of possible sequences of red and black cards, - $\binom{52}{26}$: The number of ways to arrange 26 red cards and 26 black cards.

0.0.22 Step 1: Doubling Rate W^*:

The **doubling rate** (or value gain) is defined as: $W^* = \log_2 S^*_{52}$.

Substitute $S^*_{52} = \frac{2^{52}}{\binom{52}{26}}$: $W^* = \log_2 \left(\frac{2^{52}}{\binom{52}{26}} \right)$.

Using logarithmic rules: $W^* = 52 - \log_2 \binom{52}{26}$.

0.0.23 Step 2: Calculate $\binom{52}{26}$:

The binomial coefficient is: $\binom{52}{26} = \frac{52!}{26! \cdot 26!}$.

Using **Stirling's approximation** for factorials: $n! \approx \sqrt{2\pi n} \left(\frac{n}{e} \right)^n$,

for large n: $\binom{52}{26} \approx \frac{\sqrt{2\pi \cdot 52} \left(\frac{52}{e} \right)^{52}}{\left[\sqrt{2\pi \cdot 26} \left(\frac{26}{e} \right)^{26} \right]^2}$.

Simplify: $\binom{52}{26} \approx \frac{\sqrt{52} \cdot \left(\frac{52}{e} \right)^{52}}{2\pi \cdot 26 \cdot \left(\frac{26}{e} \right)^{52}}$.

$\binom{52}{26} \approx \frac{\sqrt{52}}{2\pi \cdot 26} \cdot \left(\frac{52}{26} \right)^{52}$.

$\binom{52}{26} \approx \frac{\sqrt{52}}{2\pi \cdot 26} \cdot 2^{52}$.

0.0.24 Step 3: Simplify W^*:

Substitute into: $W^* = 52 - \log_2 \binom{52}{26}$.

Using the approximation for $\binom{52}{26}$: $\log_2 \binom{52}{26} \approx \log_2 \left(\frac{\sqrt{52}}{2\pi \cdot 26} \cdot 2^{52} \right)$.

Expand: $\log_2 \binom{52}{26} \approx \log_2 (2^{52}) + \log_2 \left(\frac{\sqrt{52}}{2\pi \cdot 26} \right)$.

$\log_2 \binom{52}{26} \approx 52 + \log_2 \left(\frac{\sqrt{52}}{2\pi \cdot 26} \right)$.

Thus: $W^* \approx 52 - \left[52 + \log_2 \left(\frac{\sqrt{52}}{2\pi \cdot 26} \right) \right]$.

$W^* \approx -\log_2 \left(\frac{\sqrt{52}}{2\pi \cdot 26} \right)$.

0.0.25 Numerical Calculation:

1. $\sqrt{52} \approx 7.211$,
2. $2\pi \cdot 26 \approx 163.36$,
3. $\frac{\sqrt{52}}{163.36} \approx 0.0441$.

$\log_2(0.0441) \approx -9.08$.

Thus: $W^* \approx 9.08\,\text{bits}$.

0.0.26 Conclusion:

The doubling rate W^* for the given scenario is 9.08 bits, as the logarithmic calculations align with the expected value. The earlier discrepancy likely arose from insufficient precision in approximating the binomial coefficient.

[ ]:

```

```