Network Optimization by Randomization
Summer Semester 2011
TU Berlin

Lecture 6

So Far...

• We start with some "random" experiment
• ...which we formalize into a "working space", i.e., a probability space \((\Omega, \mathcal{F}, P)\) with
  1. The set axioms on \(\mathcal{F}\)
  2. The measure (probability) axioms on \(P\)
• We define the event in which we are interested into
• ...and which must be an element of \(\mathcal{F}\)
• ...and whose probability can be computed by manipulating
  - the axioms
  - the independence of events (when available), to compute probabilities of intersections
  - ...or conditional probabilities (especially if the experiment can be decomposed in a series of sub-experiments)
• Various useful formulas
  - equalities (inclusion-exclusion, total probability, Bayes)
  - inequalities, when that's the best we could do ...

Random Variables

• Consider an experiment and the probability space \((\Omega, \mathcal{F}, P)\)
  - \(\Omega\) is the set of all possible outcomes (or elementary events)
  - \(\mathcal{F}\) is a set of events (i.e., subsets of outcomes)
  - \(P\) the probability measure of the events' likelihood
• Sometimes we are interested
  - not in the outcomes themselves
  - but in some functions of the outcomes

Definition (Incomplete): Random variables are real-valued functions on the sample space, i.e.,
\[ X : \Omega \to \mathbb{R} \]

Example 1

• Experiment: tossing three fair coins
\[
\Omega = \{H HH, HHT, HTT, THH, THT, TTH, TTT\} \quad F = \{\Omega\} \\
\mathbb{P}(\varnothing) = 0 \quad \forall \varnothing \in \Omega
\]
• We may simply be interested
  - in the number of "Heads" (i.e., a function of all the outcomes)
  - as opposed to some special events (e.g., 2nd coin is "Heads")
• The associated random variable \(X : \Omega \to \mathbb{R}\) takes the values
\[ X(HHH) = 3, X(HHT) = 2, X(HTT) = 2, \ldots, X(HTT) = 0 \]
• Of course, we are "always" interested in probabilities, e.g., the probability of getting a single "Heads", i.e.,
\[ P(X = 1) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\}) \]
• ...which is naturally
\[ P(X = 1) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\}) \]

Example 1 (contd.)

• ...which brings us back to computing probabilities of events
\[ P(X = 1) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\}) \]
\[ = \mathbb{P}(\{HHT, TTH, TTH\}) \]
\[ = \frac{3}{8} \]
• What has enabled this computation?
  - The assignment of probabilities (to all outcomes)
  - The assumption that \(X = \mathbb{P}\), which implies in particular that the event \(\{HHT, TTH, TTH\}\) is "measurable", i.e., it belongs to \(\mathcal{F}\)
• Note that we could also measure "the probability of getting a single "Heads" for a much more restricted probability spaces, e.g.,
\[
\Omega = \{H HH, HHT, HTT, THH, THT, TTH, TTT\} \quad F = \{\Omega, \emptyset, \{HHT\}, \{TTT\}\}
\]
\[ \mathbb{P}(\{A\}) = 0 \]
The critical fact, again, is that \(A\) is measurable, i.e. \(A \in \mathcal{F}\)
Example 2

- Experiment: rolling two fair dice
  \[ \Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \]
  \[ \mathcal{F} = \sigma \{\Omega\} \text{ where } \forall \omega \in \Omega \]

- We may be interested in:
  - whether the sum of the dice is 10
  - … as opposed to some special events (e.g., "doubles")
- We can introduce a random variable \( X : \Omega \rightarrow \{2, 3, \ldots, 12\} \) which counts the sum of the dice
- The probability of what we care (\( X = 10 \)) can be computed as
  \[ P(X = 10) = \frac{1}{12} \]

Random Variables vs. Events

- What Examples 1 & 2 & 3 have in common is that
  - Targeted probabilities (related to r.v.'s) reduce to event probabilities
  - … and these events must belong to \( \mathcal{F} \)
  - … in order for the probabilities to be computable
- … and this can be achieved by enforcing that r.v.'s are measurable functions

Example 3

- Same previous experiment but we are now interested in:
  - whether the sum of the dice is at least 10
  - … as opposed to only:
- Although what we care now is different from what we previously did, we introduce the same random variable \( X : \Omega \rightarrow \{2, 3, \ldots, 12\} \)
  - counting the sum of the dice
- The probability of what we care is
  \[ P(X = 10) = \frac{1}{6} \]
- One may also want to compute
  \[ P(X \text{ is even}) = \frac{1}{2} \]

Example 4

- Experiment: tossing a coin until first "Heads"; tosses are independent, and \( P(\text{Heads}) = p, 0 < p < 1 \)
  \[ \Omega = \{H, TH, TTH, TTTTH, \ldots\} \]
  \[ \mathcal{F} = 2^\Omega \]

- We may be interested in the number of tosses until first "Heads", i.e., until the experiment stops
- The associated random variable is \( X : \Omega \rightarrow \{1, 2, 3, \ldots\} \), and the probabilities of what we care are
  \[ P(X = 4) = P(\{w \in \Omega : X(w) = 4\}) = \frac{2p(1-p)^3}{1-p} \]
  \[ = (1-p)^3, \forall \omega \geq 1 \]

- Note: herein the r.v. encodes the elementary events

Example 5

- Same previous experiment, except that if "Heads" does not show up in the first toss, it will never do
  \[ \Omega = \{H, TH, TTH, TTTTH, \ldots\} \]
  \[ \mathcal{F} = 2^\Omega \]

- Note: \( \{TT, TTT, TTTT, \ldots\} \) are not elementary events

- The associated random variable is \( X : \Omega \rightarrow \{1, \infty\} \), and the probabilities of what we care are
  \[ P(X = 1) = P(\{w \in \Omega : X(w) = 1\}) = \frac{p}{1-p} \]
  \[ \text{and} \]
  \[ P(X = \infty) = P(\{w \in \Omega : X(w) = \infty\}) = \frac{p}{1-p} \]
Common Examples of (Discrete) Random Variables

1. **Uniform R.V.** encodes an experiment by a finite number of values
   \[ X : \Omega \rightarrow \{x_1, x_2, \ldots, x_n\} \]
   \[ X(\omega) = x_i \in \Omega, \forall \omega \in \Omega, \text{ and some mapping } f \]
   such that the probabilities are
   \[ P(X = x_i) = \frac{1}{n}, \forall i. \text{ Use notation } X : (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) \]

   Simplest example is to encode the equally likely elementary events of some experiment

   Another example: encode the sum of two dice in an experiment where each sum is equally likely
   \[ X(11) = 2, X(12) = 3, X(21) = 3, \ldots, X(66) = 12 \]

   Is such an experiment possible if the sequence of throws is independent?

Common Examples (contd.)

2. **Bernoulli R.V.** encodes (sub-)experiments with only two outcomes:
   \[ X : \Omega \rightarrow \{0, 1\} \]
   \[ X(\omega) = 1, X(\omega) = 0 \]
   \[ P(X = 1) = p, P(X = 0) = 1 - p \]

   For the experiment "Toss a coin once" with \( \Omega = \{H, T\} \), then the r.v. is \( X(T) = 0, X(H) = 1 \) with some probabilities

   But how to define Bernoulli r.v.'s for the experiment "Toss a coin until first H" with \( \Omega = \{H, TH, TT, TTH, TTT, \ldots\} \)?

   For the first toss, the Bernoulli r.v. \( X_1 : \Omega \rightarrow \{0, 1\} \)
   \[ X_1(H) = 1, \text{ and } X_1(T^k) = 0, \forall k \neq 1 \]

   For the second toss, the Bernoulli r.v. \( X_2 : \Omega \rightarrow \{0, 1\} \)
   \[ X_2(H) = 1, \text{ and } X_2(T^kH) = 0, \forall k \neq 1 \]

   More natural definitions by considering the alternative sample space \( \Omega = \{H, T\} \)

Common Examples (contd.)

3. **Binomial R.V.** encodes the "number of successes" in an experiment consisting of a sequence of \( n \) independent sub-experiments with two outcomes
   \[ X : \Omega \rightarrow \mathbb{R}_+ \]
   \[ X(\omega) = 0 \text{ if } \omega \text{ is a sequence of } n \text{ successes in } s \]

   If \( p \) is the probability of success in each sub-experiment, then the probabilities are
   \[ P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \forall 0 \leq k \leq n \]

   Note that a Binomial r.v. \( X \sim \text{Bin}(n, p) \) can be modeled as a sum of Bernoulli r.v.'s \( X = \sum_{i=1}^{n} X_i \)

   For the experiment "Toss a fair coin (independently) for 100 times" the Binomial r.v. is
   \[ X \sim \text{Bin}(100, \frac{1}{2}) \]

   Note: the sample space can be more general (e.g., balls-and-bins)

Distribution Functions

- So far ...
  - random variables are measurable real functions \( X : \Omega \rightarrow \mathbb{R} \)
  - and conveniently encode events of interest
- It is further convenient to encode the random variables themselves using also real functions, but on the real line (r.v.'s co-domain)

- Definition: The distribution function \( F : \mathbb{R} \rightarrow \mathbb{R} \) of a r.v. \( X \) is defined as
  \[ F(x) = P(X \leq x), \forall x \in \mathbb{R} \]

- Some useful properties
  1. \( F(x) \) is non-decreasing in \( x \)
  2. \( F(-\infty) = 0, F(\infty) = 1 \)
  3. \( P(a < X \leq b) = F(b) - F(a) \)

From the Experiment -> Probability Space -> Formulating the Probability of Interest

- The experiment can be restated as follows: we throw (uniformly and independently) \( n \) balls into \( m \) bins
- The probability space is
  \[ \Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}, |\Omega| = m^n \]
  \[ \mathcal{F} = \mathcal{P}(\Omega) \]

- We now introduce the random variables \( X_1 : \Omega \rightarrow \{0, 1, \ldots, m-1\} \)
  where \( X_i \) denotes the # of collisions of ball \( i \) \( i = 1, 2, \ldots, \)
- The probability of interest can be written as
  \[ P(\text{collisions}) = P(X_1 > 0) \land (X_2 > 1) \land \ldots \land (X_m > 1) \leq \sum_{i=1}^{m} P(X_i > 1) \]

- after using the union bound (for some events \( A_1^\prime \))

Killer App. Hashing

- Problem (from last time): How many (randomly chosen) keys can be hashed (uniformly and independently) before the probability of a collision exceeds some value (e.g., \( 1/2 \))?}

- More formally, we are given
  - A (very large) set of keys \( U = \{x_i : i = 1, \ldots, m\} \)
  - A (hash) table \( T = \{y_j : j = 1, 2, \ldots, n\} \)
  - A (hash) function \( f : U \rightarrow T \) which distributes the keys evenly, i.e.,
    \[ y_j \in T, \forall x_i \in U : f(x_i) = y_j \]
  - A subset of keys \( G \) with \( |G| = a \), which are uniformly and independently selected from \( U \)

- Question: What is the maximum value of \( a \) such that ... ?

  \[ P(\text{collisions}) \leq \frac{1}{m}, \text{ i.e., very small} \]
**Solution**

- Denote the events $1_j := \{\text{ball 1 in bin } j, j = 1, 2, \ldots, n\}$
- Let's first calculate
  \[
  P(X_1 = 1) = \sum_j P(1_j) P(2_j)
  \]
  \[
  = \sum_j \left(1 - \frac{1}{n}\right)^{n-1} \cdot \frac{1}{n}
  \]
  "See" the conditional probability space
  \[
  = \left(1 - \frac{1}{n}\right)^{n-1}
  \]
- We can now estimate
  \[
  P(\text{collisions}) \leq \sum_j P(X_j > 1) = \sum_j \left(1 - \left(1 - \frac{1}{n}\right)^{n-1}\right)
  \]
  \[
  \leq \sum_j \left(1 - 1 + \frac{1}{n+1}\right) = \sum_j \left(1 - 1 + \frac{1}{n+1}\right)
  \]
- Therefore
  \[
  a_n = \exists k \text{ such that } P(\text{collisions}) \leq \frac{1}{a_n}
  \]

**Killer App. Load Balancing**

- **Problem:** Consider $n$ jobs which are to be randomly (uniformly, independently) assigned to $n$ processors. What is the minimum value $k$ such that the probability that the maximum number of jobs exceeds $k$ is very small.

- The experiment can be restated as follows: we throw (uniformly and independently) $n$ balls into $n$ bins

- The probability space is

  \[
  \Omega = \{\pi_1, \pi_2, \ldots, \pi_n : \pi_i \in \{1, 2, \ldots, n\}\}, \left|\Omega\right| = n^n
  \]

  \[
  F = \frac{1}{n^n}
  \]

- We now introduce the random variables $X_\ell : \Omega \rightarrow \{0, 1, \ldots, n\}$

  where $X_\ell$ denotes the number of balls in bin $\ell$

- We look for the minimum $k$ such that

  \[
  P(\max X_\ell \geq k) \leq \frac{1}{a_n}
  \]