Lecture 17

Last time before review
  • Poisson approximation

Today
  • Discuss midterm
  • Review Poisson approximation
  • Application: max load
  • Application: Coupon Collector
Let $X$ and $Y$ be random variables that take values in $\mathbb{N}$.

Which of the following are NOT well defined?

A. $\Pr[X]$

B. $\Pr[X = Y]$

C. $\Pr[X \cap Y]$

D. $\Pr[X \text{ is a geometric RV with parameter } p]$

E. $\mathbb{E}[\mathbb{E}[X]]$

F. $\mathbb{E}[X = 0]$

Make sure your objects are of the right type!
Choose all that apply:

Random variables $X$ and $Y$ are independent if

A. $\Pr[X \cap Y] = \Pr[X] \cdot \Pr[Y]$

B. $\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \cdot \Pr[Y = y]$

C. $\Pr[(X = x) \cap (Y = y)] = \Pr[X = x] \cdot \Pr[Y = y]$ for all values $x$ and $y$.

D. $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

E. $\text{cov}(X, Y) = 0$
Review question

Let $X$ be $\text{Bin}(10, p)$ and $Y$ be uniform over $\{0, 1, \ldots, 10\}$. Compute $\Pr[X > Y]$. 
A Poisson random variable with parameter $\mu$ is given by the following distribution on $j = 0, 1, 2, \ldots$

$$\Pr[X = j] = \frac{e^{-\mu} \mu^j}{j!}$$

- $\mathbb{E}[X] = \text{var}[X] = \mu$
- Sum of two Poisson RVs is a Poisson RV
- A Poisson RV satisfies Chernoff-type bounds.
- Poisson distribution is a limit of binomial distribution
Poisson approximation

- The Balls-and-Bins model has dependences.
- The Poisson approximation gets rid of dependencies.

- $m$ balls into $n$ bins u.i.r.

\[ \begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & \ldots & m-1 & m \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{array} \]

For $i \in [n]$, let

\begin{align*}
\text{(real world)} & \quad X_i^{(m)} = \# \text{ of balls in bin } i \\
\text{(Poisson world)} & \quad Y_i^{(m)} \sim \text{Poisson}(\mu), \text{ where } \mu = \frac{m}{n} \text{ and } \\
& \quad Y_i^{(m)} \text{ are mutually independent.}
\end{align*}

- If we condition the Poisson distribution on producing exactly $k$ balls, then it’s the same as the distribution resulting from throwing $k$ balls into $n$ bins.
Approximating a function of the loads of the bins

Poisson Approximation Theorem

Let $f(x_1, ..., x_n) \geq 0$ for all $x_1, ..., x_n \in \{0,1,2, ... \}$. Then

- **Fact (Stirling's formula):** $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

- **Bounds for all $n \in \mathbb{N}$:** $\sqrt{2\pi} \sqrt{n} \left(\frac{n}{e}\right)^n \leq n! \leq e \sqrt{n} \left(\frac{n}{e}\right)^n$
Approximating a function of the loads of the bins

<table>
<thead>
<tr>
<th>Poisson Approximation Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let $f(x_1, ..., x_n) \geq 0$ for all $x_1, ..., x_n \in {0,1,2, ... }$. Then</td>
</tr>
<tr>
<td>$\mathbb{E} \left[ f \left( X_1^{(m)}, ..., X_n^{(m)} \right) \right] \leq e^{\sqrt{m}} \cdot \mathbb{E} \left[ f \left( Y_1^{(m)}, ..., Y_n^{(m)} \right) \right]$.</td>
</tr>
</tbody>
</table>

- **Poisson case:** # of balls in each bin is independent Poisson($\frac{m}{n}$)
- **Corollary.** Any event that has probability $p$ in the Poisson case has probability $\leq p \cdot e^{\sqrt{m}}$ in the exact case.

**Proof:** Let $X$ be the indicator for that event.

Then $\mathbb{E}[X]$ is the probability that event occurs.

- **Improvements to Theorem and Corollary**

If $\mathbb{E} \left[ f \left( X_1^{(m)}, ..., X_n^{(m)} \right) \right]$ is monotonically nonincreasing (or nondecreasing) in $m$, then $e^{\sqrt{m}}$ can be changed to 2.
Application: Max Load

$n$ balls into $n$ bins

- **Before (by Chernoff):** \( \Pr \left[ \text{MaxLoad} > \frac{3 \ln n}{\ln \ln n} \right] \leq \frac{1}{n} \) for s.l. $n$

- **Theorem.** \( \Pr \left[ \text{MaxLoad} < \frac{\ln n}{\ln \ln n} \right] \leq \frac{1}{n} \) for s.l. $n$

**Proof:** Let $M = \frac{\ln n}{\ln \ln n}$
Application: Coupon Collector

\( X = \# \text{ of coupons observed before obtaining 1 of each of } n \text{ types} \)

- **Before:** \( \mathbb{E}[X] = \) and \( \Pr[X > n \ln n + cn] \leq e^{-c} \ \forall c > 0 \)
- **Review:** \( \Pr[\text{not obtaining coupon } i \text{ in } n \ln n + cn \text{ steps}] \) is

- **Theorem.** \( \Pr[X > n \ln n + cn] \leq 2(1 - e^{-1.5 \cdot e^{-c}}) \ \forall c > 0, \text{ s.l. } n \)
- **MU:** \( \lim_{n \to \infty} \Pr[X > n \ln n + cn] = 1 - e^{-e^{-c}} \ \forall c > 0 \)
Application: Coupon Collector

- **Theorem.** \( \Pr[X > n \ln n + cn] \leq 2(1 - e^{1.5 \cdot e^{-c}}) \forall c > 0, \text{s.l. } n \)

**Proof:** Balls and bins view

- \( n \) bins
- \( X = \# \text{ balls thrown before all bins nonempty} \)
- Consider \( m = n(\ln n + c) \) balls.
- Let \( B = \text{ event that there is an empty bin.} \)

\[
\Pr[X > n \ln n + cn] = \Pr[B]
\]

**Idea:** Use Poisson approximation:

- # balls in each bin is Poisson(\( \mu \)) with \( \mu = \ln n + c \)
- \( E_i = \text{ event that bin } i \text{ is empty. Then} \)

\[
\Pr[E_i] = e^{-\mu} = e^{-(\ln n + c)} = \frac{e^{-c}}{n}
\]
Theorem. \( \Pr[X > n \ln n + cn] \leq 2(1 - e^{-1.5 \cdot e^{-c}}) \) \( \forall c > 0 \), s.l. \( n \)

Proof: \( n \) bins and \( m = n(\ln n + c) \) balls.

- Let \( B \) = event that there is an empty bin: \( \Pr[X > n \ln n + cn] = \Pr[B] \)
- Poisson approximation: \# balls in each bin is Poisson(\( \mu \)), \( \mu = \ln n + c \)
- \( E_i = \) event that bin \( i \) is empty. Then \( \Pr[E_i] = \frac{e^{-c}}{n} \)

\[ e^{-1.5x} = e^{-x-0.5x} \leq e^{-x-x^2} \leq 1 - x \leq e^{-x} \]