CSE 312

Foundations of Computing II

Lecture 19: Concentration Wrap-Up + Introduction to Continuous Random Variables

Stefano Tessaro
tessaro@cs.washington.edu
Today – Two things

• Wrap-up concentration / tail inequalities [very quick]
• Introduction to continuous (i.e., non-discrete) random variables

Also: HW5 is online + Naïve Bayes due!
• Extra office hour today with Kushal

Midterm results will come today.
Concentration / tail bounds – A guided tour

https://us.edstem.org/courses/125/discussion/8100
# Tail Bounds – Summary

**Goal:** We need to compute $\mathbb{P}(X > t)$ for $t > \mathbb{E}(X)$

<table>
<thead>
<tr>
<th>If we know ...</th>
<th>... we use ...</th>
<th>... to obtain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu = \mathbb{E}(X)$</td>
<td>Markov</td>
<td>$\mathbb{P}(X &gt; t) &lt; \frac{\mu}{t}$</td>
</tr>
<tr>
<td>$\mu = \mathbb{E}(X)$ and $\sigma^2 = \text{Var}(X)$</td>
<td>Chebyshev</td>
<td>$\mathbb{P}(X &gt; t) &lt; \frac{\text{Var}(X)}{(t-\mu)^2}$</td>
</tr>
<tr>
<td>$X = X_1 + \cdots + X_n$, sum of indep. RVs in $[0,1], \mu = \mathbb{E}(X)$</td>
<td>Chernoff</td>
<td>$\mathbb{P}(X &gt; t) &lt; e^{-\frac{\epsilon^2}{2+\epsilon}}$, where $\epsilon = \frac{t-\mu}{\mu}$</td>
</tr>
</tbody>
</table>

Chernoff usually wins when $\mu$ grows as a function of $n$
Goals: We need to compute $\mathbb{P}(\{|X - \mu| > \epsilon \mu\})$ for $\epsilon > 0$ and $\mu = \mathbb{E}(X)$

<table>
<thead>
<tr>
<th>If we know ...</th>
<th>... we use ...</th>
<th>... to obtain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$ only</td>
<td>Out of luck!</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2 = \text{Var}(X)$</td>
<td>Chebyshev</td>
<td>$\mathbb{P}({</td>
</tr>
<tr>
<td>$X = X_1 + \cdots + X_n$, sum of indep. RVs in $[0,1]$</td>
<td>Chernoff</td>
<td>$\mathbb{P}({</td>
</tr>
</tbody>
</table>

Chernoff usually wins when $\mu$ grows as a function of $n$
Sampling Theorem – Recap

• **M** individuals, a fraction \( p \in [0,1] \) is in favor of CSE313
• **Goal:** Produce good estimate \( \hat{P} \) of \( p \)
• **Idea:**
  – Ask \( n < M \) randomly selected individuals whether they want CSE313
  – Responses are Bernoulli \( X_1, ..., X_n \) with parameter \( p \)
  – Let \( \hat{P} = \frac{1}{n} \sum_{i=1}^{n} X_i \rightarrow \mathbb{E}(\hat{P}) = p \)

• **Sampling theorem:** If \( n \geq \ln(1/\delta) \frac{2+\theta}{\theta^2} \), then \( \mathbb{P}(|\hat{P} - p| \leq \theta) \geq 1 - \delta \)
  – \( \theta = \) how good is the estimate
  – \( \delta = \) probability we fail to provide good estimate
Sampling Theorem – Proof

\[
\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \mathbb{E}(\hat{p}) = p \quad \mathbb{P}(X_i = 1) = p
\]

\[
\mathbb{P}(|\hat{P} - p| > \theta) = \mathbb{P}(|n\hat{P} - np| > n\theta)
\]

\[
= \mathbb{P}(|\sum_{i}^{n} X_i - np| > n\theta)
\]

\[
= \mathbb{P}\left(|\sum_{i}^{n} X_i - np| > np \frac{\theta}{p}\right)
\]

\[
< 2 \exp\left(-\frac{\theta^2/p^2}{2 + \theta/p} pn\right)
\]

\[
= 2 \exp\left(-\frac{\theta^2}{2p + \theta} n\right) \leq 2 \exp\left(-\frac{\theta^2}{2 + \theta} n\right)
\]

Need to rephrase in terms of relative error to use Chernoff Bound with \(\epsilon = \frac{\theta}{p}\) and \(\mu = np\)!

Remove dependency on \(p\)
Sampling Theorem – Proof (cont’d)

We have proved:

\[ \mathbb{P}(|\hat{P} - p| > \theta) < 2 \exp\left(-\frac{\theta^2}{2 + \theta n}\right) \]

We have \( 2 \exp\left(-\frac{\theta^2}{2+\theta} n\right) \leq \delta \) if (and only if)

\[ n \geq \ln(1/\delta) \frac{2 + \theta}{\theta^2} \]
Bottom line: Often we want to model experiments where the outcome is not discrete.
Example – Lightning Strike

Lightning strikes a pole within a one-minute time frame

- $T = $ time of lightning strike
- Every time within $[0,1]$ is equally likely
  - Time measured with infinitesimal precision.

$$T = 0.71237131931129576 \ldots$$
Lightning strikes a pole within a one-minute time frame

- $T$ = time of lightning strike
- Every point in time within $[0,1]$ is equally likely

\[
\mathbb{P}(T \geq 0.5) = \frac{1}{2}
\]
Lightning strikes a pole within a one-minute time frame

- \( T \) = time of lightning strike
- Every point in time within \([0,1]\) is equally likely

\[ P(0.2 \leq T \leq 0.5) = 0.5 - 0.2 = 0.3 \]
Lightning strikes a pole within a one-minute time frame

- $T$ = time of lightning strike
- Every point in time within $[0,1]$ is equally likely

$\mathbb{P}(T = 0.5) = 0$
Bottom line

• This gives rise to a different type of random variable
• $\mathbb{P}(T = x) = 0$ for all $x \in [0,1]$
• Yet, somehow we want
  – $\mathbb{P}(T \in [0,1]) = 1$
  – $\mathbb{P}(T \in [a, b]) = b - a$
  – ...
• How do we model the behavior of $T$?
Definition. A continuous random variable $X$ is defined by a probability density function (or simply, “density”) $f_X: \mathbb{R} \to \mathbb{R}$ such that

- **Non-negativity:** $f_X(x) \geq 0$ for all $x \in \mathbb{R}$
- **Normalization:** $\int_{-\infty}^{+\infty} f_X(x) \, dx = 1$
PDF of Uniform RV

\( T \sim \text{Unif}(0,1) \)

\[
f_T(x) = \begin{cases} 
1, & x \in [0,1] \\
0, & x \notin [0,1]
\end{cases}
\]

\[
\int_{-\infty}^{+\infty} f_T(x) \, dx = \int_{0}^{1} f_T(x) \, dx = 1 \cdot 1 = 1
\]

Density ≠ Probability

\( f_T(0.5) = 1 \quad \mathbb{P}(T = 0.5) = 0 \)
Probability of Event

Definition. \( \mathbb{P}(X \in S) = \int_S f_X(x) \, dx \)

Example. \( T \sim \text{Unif}(0,1) \)

\[ f_T(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases} \]

\[ \mathbb{P}(T \in [a,b]) = \int_a^b f_T(x) \, dx = b - a \]
PDF of Uniform RV

\[ X \sim \text{Unif}(0,0.5) \]

\[
f_X(x) = \begin{cases} 
2, & x \in [0,0.5] \\
0, & x \notin [0,0.5] 
\end{cases}
\]

\[
\int_{-\infty}^{+\infty} f_X(x) \, dx = \int_{0}^{1} f_X(x) \, dx = 2 \cdot 0.5 = 1
\]

Intuition: \( \mathbb{P}(X \in [x - \epsilon, x + \epsilon]) \approx f_X(x) \cdot \epsilon \) for small \( \epsilon \)
**Definition.** The **cumulative distribution function (cdf)** of $X$ is defined as

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x) \, dx$$

Therefore: $\mathbb{P}(X \in [a, b]) = F(b) - F(a)$
Example. $T \sim \text{Unif}(0,1)$

PDF:

$$f_T(x) = \begin{cases} 
1, & x \in [0,1] \\
0, & x \notin [0,1] 
\end{cases}$$

CDF:

$$F_T(x) = \begin{cases} 
0, & x \leq 0 \\
x, & 0 \leq x \leq 1 \\
1, & 1 \leq x 
\end{cases}$$
Properties of Random Variables

• Quantities of random variables – $\mathbb{E}(X)$, $\text{Var}(X)$, ... – generalize naturally from discrete to continuous RVs
  – Usually have the same properties

• Basic idea: $\sum \rightarrow \int$
### Expectation of a Continuous RV

**Definition.** The **expected value** of a continuous RV $X$ is defined as

$$
\mathbb{E}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot x \, dx
$$

**Fact.** \( \mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c \)

**Definition.** The **variance** of a continuous RV $X$ is defined as

$$
\text{Var}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot (x - \mathbb{E}(X))^2 \, dx = \mathbb{E}(X^2) - \mathbb{E}(X)^2
$$
Expectation of a Continuous RV

**Definition.**

\[
\mathbb{E}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot x \, dx
\]

**Example.** \( T \sim \text{Unif}(0,1) \)

\[
f_T(x) = \begin{cases} 
1, & x \in [0,1] \\
0, & x \notin [0,1]
\end{cases}
\]

\[
f_T(x) \cdot x = \begin{cases} 
x, & x \in [0,1] \\
0, & x \notin [0,1]
\end{cases}
\]

\[
\mathbb{E}(T) = \frac{1}{2} \cdot 1^2 = \frac{1}{2}
\]

Area of triangle