Overview

- Markov’s Inequality
- Chebyshev’s Inequality
- Chernoff’s Inequality
Why Are Inequalities Useful?

We may not know the true form of a probability distribution

- Opinion polls
  The 2016 General Election Exit Poll was conducted exclusively on behalf of *The Irish Times* by Ipsos MRBI, among a national sample of 5,260 voters at 200 polling stations throughout all constituencies in the Republic of Ireland.

  Voters were randomly selected to self-complete a mock ballot paper on exiting the polling station. The accuracy level is estimated to be approximately plus or minus 1.2 per cent.

- Stock market data
- Weather tomorrow

But we may know some of its properties

- Mean
- Variance
- Non-negativity

Inequalities let us say something about the probability distribution in such cases, although often imprecise. They are also very important for looking at what happens as we collect more and more measurements.
Markov’s Inequality

Often we want to know:

What is the probability that the value of r.v. $X$ is “far” from its mean?

A generic answer for non-negative $X$ is Markov's inequality. Say $X$ is a non-negative random variable. Then:

$$P(X \geq a) \leq \frac{E(X)}{a} \text{ for all } a > 0$$

Proof:

• Let indicator $I_a(X) = 1$ if $X \geq a$ and $I_a(X) = 0$ otherwise. Then $aI_a(X) \leq a \leq X$ i.e. $I_a(X) \leq \frac{X}{a}$.

• $E(I_a(X)) \leq E\left(\frac{X}{a}\right) = \frac{E(X)}{a}$

• $E(I_a(X)) = P(X \geq a) \leq \frac{E(X)}{a}$
Markov’s Inequality

Andrey Andreyevich Markov (1856-1922) was a Russian mathematician

- Markov’s inequality is named after him
- Also Markov Chains, used e.g. in Google’s PageRank algorithm
Markov’s Inequality

Example: Roll 6-sided dice.

- Mean is $E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$
- Markov inequality: $P(X \geq 5) \leq \frac{3.5}{5} = 0.7$. Exact: $P(X = 5) + P(X = 6) = \frac{1}{6} + \frac{1}{6} = 0.33$
- So a loose bound, but it made no assumptions about the form of distribution.
Markov’s Inequality

Example: Distribution of number $X$ of Facebook friends.

- Mean is $E(X) = 190$ (!)
- Markov inequality: $P(X \geq 500) \leq \frac{190}{500} = 0.38$. From plot, $P(X \geq 500) \approx 0.1$.
- Markov inequality: $P(X \geq 190) \leq \frac{190}{190} = 1$, non-informative. From plot, $P(X \geq 190) \approx 0.3$.

source: http://arxiv.org/abs/1111.4503
Markov’s Inequality

Example: IQ in Ireland

- Mean is $E(X) = 92$. Score 110-119 = “high average”, 120-129 = “superior”.
- Markov inequality: $\Pr(X \geq 110) \leq \frac{92}{110} = 0.83$. From data, $\Pr(X \geq 110) \approx 0.11$.
- Markov inequality: $\Pr(X \geq 120) \leq \frac{92}{120} = 0.76$. From data, $\Pr(X \geq 120) \approx 0.029$. 

<table>
<thead>
<tr>
<th>IQ Score</th>
<th>P(IQ ≥ x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>60</td>
<td>0.1</td>
</tr>
<tr>
<td>80</td>
<td>0.2</td>
</tr>
<tr>
<td>100</td>
<td>0.3</td>
</tr>
<tr>
<td>120</td>
<td>0.4</td>
</tr>
<tr>
<td>140</td>
<td>1</td>
</tr>
</tbody>
</table>
Chebyshev’s Inequality

Suppose $X$ is a random variable with mean $E(X) = \mu$ and variance $\text{var}(X) = \sigma^2$. Then

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \text{ for all } k > 0$$

Proof:

• Since $(X - \mu)^2$ is a non-negative random variable we can apply Markov’s inequality with $a = k^2$ to get

$$P((X - \mu)^2 \geq k^2) \leq \frac{E((X - \mu)^2)}{k^2} = \frac{\sigma^2}{k^2}$$

• Note that $(X - \mu)^2 \geq k^2 \iff |X - \mu| \geq k$, so

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$
Chebyshev’s Inequality

Pafnuty Lvovich Chebyshev (1821 - 1894) also a Russian mathematician

- Chebyshev’s inequality was in fact first formulated by French mathematician Jules Bienaymé without proof, then proved by Chebyshev 14 years later.
- Markov was a graduate student of Chebyshev (also Aleksandr Lyapunov, but that’s another days work)
**Chebyshev’s Inequality**

Chebyshev’s inequality links the “spread” of values of a random variable around its mean to the variance $\sigma^2$:

- Applying Chebyshev’s inequality $P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$ with $k = n\sigma$ gives:

$$P(|X - \mu| \geq n\sigma) \leq \frac{1}{n^2}$$

- With $n = 3$ then $P(|X - \mu| \geq 3\sigma) \leq \frac{1}{9} = 0.11$.
- This holds even when distribution is not Gaussian, so can be quite handy (if conservative).
Chebyshev’s Inequality

Example: Roll 6-sided dice.

- Mean is $E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$
- Variance is $Var(X) = E[X^2] - E[X]^2$.
  
  $E[X^2] = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6} \approx 15.17$
  
  $Var(X) = 15.17 - 3.5^2 \approx 2.9$
- Chebyshev inequality: $P(\left|X - 3.5\right| \geq 2.5) \leq \frac{2.9}{2.5^2} = 0.46$.
- Exact: $P(\left|X - 3.5\right| \geq 2.5) = P(X = 1) + P(X = 6) = \frac{1}{6} + \frac{1}{6} = 0.33$
- A loose bound, but use of variance in Chebyshev inequality can improve accuracy cf Markov inequality.
Chebyshev’s Inequality

Example: IQ in Ireland

- Mean is $E(X) = 92$, variance is $\sigma^2 = 225$.
- Chebyshev inequality: $P(|X - 92| \geq 20) \leq \frac{225}{400} = 0.56$. From data, $P(|X - 92| \geq 20) \approx 0.18$.
- Markov inequality: $P(X \geq 112) \leq \frac{92}{112} = 0.82$. And need to add $P(X \leq 72)$ to this.
Chernoff’s Inequality

Suppose we have random variable $X$. Then

$$P(X \geq a) \leq \min_{t>0} e^{-ta} e^{\log E(e^{tX})}$$

Proof:

- $P(X \geq a) = P(e^{tX} \geq e^{ta})$ for $t > 0$.
- By Markov’s inequality

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}} = e^{-ta} E(e^{tX})$$

- This holds for all $t > 0$, so might as well choose the one that minimises it.
Chernoff’s Inequality

Herman Chernoff is a US mathematician (with Russian parents)

- Was at MIT, then Harvard.
Chernoff’s Inequality

- $P(X \geq a) \leq \min_{t>0} e^{-ta} e^{\log E(e^{tX})}$
- $E(e^{tX})$ is called the “moment generating function”
- Contains more information about the distribution than just the mean (used by Markov inequality) and variance (used by Chebyshev inequality).
Chernoff’s Inequality

Example: coin flipping:

- A fair coin lands on heads with probability $1/2$ and on tails with probability $1/2$.
- If coin is flipped 100 times, give an upper bound on the probability that it lands heads at least 60 times.
- Random variable $X_k = 1$ is heads, $X = 0$ if tails at flip $k$.

$$S = \sum_{k=1}^{100} X_k.$$ 

$$E(e^{tX_k}) = \frac{1}{2} e^{t\times1} + \frac{1}{2} e^{t\times0} = \frac{1}{2}(e^t + 1)$$

$$\log E(e^{tS}) = \log E\left(\prod_{k=1}^{100} e^{tX_k}\right) = \log \prod_{k=1}^{100} E(e^{tX_k}) = 100 \log\left(\frac{1}{2}(e^t + 1)\right)$$

- By Chernoff’s inequality, the probability of at least 55 heads is

$$P(S \geq 60) \leq \min_{t>0} e^{-60t} e^{\log E(e^{tX})} = \min_{t>0} e^{-60t} e^{100 \log\left(\frac{1}{2}(e^t + 1)\right)}$$
Chernoff’s Inequality

- $P(S \geq 60) \leq \min_{t>0} e^{-60t} \cdot e^{100 \log(\frac{1}{2}(e^t+1))}$

- Using $t = 0.4$, Chernoff’s inequality gives $P(S \geq 60) \leq 0.13$.
- Markov inequality gives $P(S \geq 60) \leq \frac{E(S)}{60} = \frac{50}{60} = 0.83$
Chernoff’s Inequality

• Let’s try it ...

```matlab
XX = []; for i = 1:20000,
    XX = [XX, sum(rand(1,100) < 0.5)];
end
[n, x] = hist(XX, 1000);
plot(x, n)
xlabel('Number of heads')
ylabel('Count')
```

• Out of 20,000 trials 592 have \( \geq 60 \) heads i.e 0.03. Cf Markov inequality value of 0.83 and Chernoff inequality value of 0.13.