ECE353: Probability and Random Processes

Lecture 17 - Concentration and Confidence Interval

Xiao Fu

School of Electrical Engineering and Computer Science
Oregon State University
E-mail: xiao.fu@oregonstate.edu
Markov’s Inequality

• **Theorem:** Consider a nonnegative RV \( X \) \((P[X < 0] = 0)\). We have

\[
P[X \geq c^2] \leq \frac{E[X]}{c^2}
\]

• **Proof:**

\[
P[X \geq c^2] = \int_{x=0}^{\infty} u(x - c^2) f_X(x) \, dx,
\]

where

\[
u(x) = \begin{cases} 
  1, & x \geq 0 \\
  0, & \text{o.w.}
\end{cases}
\]
Markov’s Inequality

- Pictorially as below, hence $f_X(x) u(x - c^2) \leq f_X(x) x / c^2$. 

\[
y = \frac{x}{c^2}
\]

\[
y = u(x - c^2)
\]

\[
f_X(x) u(x - c^2)
\]
Markov’s Inequality

• Therefore, we have

\[ P[X \geq c^2] = \int_{x=0}^{\infty} u(x - c^2) f_X(x) \, dx \leq \int_{x=0}^{\infty} \left( x/c^2 \right) f_X(x) \, dx = \frac{E[X]}{c^2} \]

• **Significance:** knowing only the expected value you get a bound on the “tail probability”!

• Also note: \( F_X(c^2) := P[X \leq c^2] = 1 - P[X \geq c^2] \) (assuming continuous RV)
\[ \implies F_X(c^2) \geq 1 - \frac{E[X]}{c^2}. \]
**Chebyshev’s Inequality**

- Consider \(X = (Y - \mu_Y)^2\), \(\mu_Y = E[Y]\). Apply Markov’s inequality to \(X\), we have

\[
P \left[ (Y - \mu_Y)^2 \geq c^2 \right] \leq \frac{E[(Y - \mu_Y)^2]}{c^2}
\]

Assuming \(c \geq 0\), we have

\[
P \left[ |Y - \mu_Y| \geq c^2 \right] \leq \frac{\text{Var}[Y]}{c^2},
\]

which is the **Chebyshev’s Inequality**.
Chebyshev’s Inequality

• Application: consider a set of RVs \( \{X_i\}_{i=1}^n \), which are uncorrelated to each other, i.e., \( \text{Cov}[X_i, X_j] = 0 \). We also have

\[
E[X_i] = \mu, \quad E[(X_i - \mu)^2] = \sigma^2, \quad \forall i.
\]

• Define \( \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \). We know that \( E[\hat{\mu}] = \mu \) and

\[
\text{Var}[\hat{\mu}] = \frac{1}{n^2} \text{Var} \left[ \sum_{i=1}^n X_i \right] = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}. \quad \text{(used uncorrelatedness)}
\]

• Applying Chebyshev’s inequality we have

\[
P[|\hat{\mu} - \mu| \geq c] \leq \frac{\text{Var}[\hat{\mu}]}{c^2} = \frac{\sigma^2}{nc^2}
\]

• Can you see how significant is this?
Convergence of Sequence of RVs

• The sequence of RVs $\{\hat{\mu}\}_{n=1}^{\infty}$ converges in probability to the constant $\mu$.

• This result is also called “the weak law of large numbers”.

• **Remark:** we have also seen that $E[(\hat{\mu} - \mu)^2] = \text{Var}[\hat{\mu}]$. Hence,

  $$E[(\hat{\mu} - \mu)^2] \to 0, \text{ when } n \to \infty.$$  

• When this happens, we say that the sequence of RVs $\{\hat{\mu}\}_{n=1}^{\infty}$ converges to the constant $\mu$ in the **mean square sense**.

• By Chebyshev’s inequality, convergence in the mean square sense implies convergence in probability.
Confidence Interval

• Let us start again with Chebyshev’s inequality

\[ P \left[ \left| Y - \mu_Y \right| \geq c^2 \right] \leq \frac{\text{Var}[Y]}{c^2}. \]

Taking \( Y = (1/n) \sum_{i=1}^{n} X_i \), \( \text{Var}[Y] = \text{Var}[X]/n = \sigma^2/n \). Therefore

\[ P \left[ \left| Y - \mu_Y \right| \geq c^2 \right] \leq \frac{\text{Var}[X]}{nc^2} \]

and

\[ P \left[ \left| Y - \mu_Y \right| < c^2 \right] \leq 1 - \alpha \]

where \( \alpha = \frac{\text{Var}[X]}{nc^2} \).

• \( P[Y - c < \mu_Y < Y + c] \geq 1 - \alpha \), where 1 - \( \alpha \) is called the confidence level and \([Y - c, Y + c]\) is called a confidence interval of length 2c.
Confidence Interval

• **Take-home point**: for a designed accuracy (determined by $c$), if you know $\text{Var}[X]$ or an upper bound of $\text{Var}[X]$, you can pick $n$, such that you attain a certain confidence level $1 - \alpha$.

• **Example** (Polling for Election). Consider a party or a candidate, denoted by $A$. We are interested in estimating $P[A]$, i.e., the probability that any voter votes for the party/candidate. Let $Z_i$ be a RV representing vote of voter $i$. Let us define an auxiliary RV $X_i$ as

$$X_i = 1(Z_i = A) = \begin{cases} 1, & Z_i = A \\ 0, & \text{o.w.} \end{cases}$$

• Note that $E[X_i] = 1 \cdot P[X_i = 1] + 0 \cdot P[X_i = 0] = P[Z_i = A] = P[A]$.

• Hence, our goal is to estimate $E[X_i]$, i.e., $\hat{P}_n[A] = \frac{1}{n} \sum_{i=1}^{n} X_i$. 

Confidence Interval

- Our goal is to estimate $E[X_i]$, i.e., $\hat{P}_n[A] = \frac{1}{n} \sum_{i=1}^{n} X_i$.

- Apply Chebyshev’s inequality:

\[ P[|\hat{P}_n[A] - P[A]| < c] \geq 1 - \frac{\text{Var}[X_i]}{nc^2}. \]

- $\text{Var}[X_i] = ?$ Note that $X_i$ is binary (0 – 1) and thus $E[X_i^2] = E[X_i] = P[A]$.

\[ \text{Var}[X_i] = E[X_i^2] - (E[X_i])^2 \\
= P[A] - (P[A])^2 \\
= P[A](1 - P[A]) \]

- Let $c = 0.01$ and $1 - \alpha = 0.95$. We wish to compute the $n$ we need. But we do not know $P[A]$. 

Confidence Interval

- We do not know $P[A]$. But we can consider the worst case, which has the largest $P[A](1 - P[A])$.

- Intuitively, what is the worst-case (most difficult to predict)?

- Let us consider

$$\frac{dp(1-p)}{dp} = 0 \implies 1 - 2p = 0, \ p = 1/2.$$ 

Hence,

$$P[|\hat{P}_n[A] - P[A]| < c] \geq 1 - \frac{1}{4nc^2}.$$ 

- To make $c = 0.01$ and $1 - \alpha = 0.95$, we need $n \geq 50,000$. Mission impossible for any poll.