MATH324 (Statistics) – Lecture Notes
McGill University

Masoud Asgharian

Winter 2019

Contents

1 Lecture 0

2 Lecture 1

3 Lecture 2

3.1 Markov's Inequality

Let X be a random variable and h be a **non-negative** function; ie:

$$h : R \to R^{+} \cup \{0\} \quad = \quad [0, \infty)$$

Suppose $E(h(x)) \; < \infty$,then for some $\lambda > 0$, we have:

$$P(h(x) \geq \lambda) \; \leq \; \frac{E[h(x)]}{\lambda} \tag{1}$$

Proof. Suppose X is a continuous random variable:

$$\begin{aligned}
E[h(x)] &= \int_x h(x) d_x(x) dx \\
&= \Big(\int_{x:h(x)\geq\lambda} h(x) f_x(x) dx + \int_{x:h(x)<\lambda} h(x) f_x(x) dx \Big) \\
&\geq \int_{x:h(x)\geq\lambda} h(x) f_x(x) dx && \underline{since}\ h \geq 0 \\
&\geq \lambda \int_{x:h(x\geq\lambda} f_x(x) dx = \lambda \; P(h(x) \geq \lambda)
\end{aligned}$$

$$\implies P(h(x) \geq \lambda) \leq \frac{E(h(x))}{\lambda}$$

The proof for the discrete case is similar. □

3.2 Tchebyshev's Inequality

Tchebyshev's Inequality is a special case of Markov's Inequality. Consider $h(x) = (x - \mu)^2$, then:

$$\begin{aligned}
P(|x - \mu| \geq \lambda) &= P((x - \mu)^2 \geq \lambda^2) \\
&\leq \frac{E[(x - \mu)^2]}{\lambda^2} && if\ E[(x - \mu)^2] < \infty
\end{aligned}$$

Let $\mu = E(X)$, then $E[(x - \mu)^2] = Var(X)$ denoted by σ_x^2. We therefore have:

$$P(|x - \mu_x| \geq \lambda) \leq \frac{\sigma_x^2}{\lambda^2} \qquad where\ \mu_x = E(x) \tag{2}$$

Now consider $\lambda = K\sigma_x$ where K is a known number. Then:

$$P(|x - \mu_x| \geq K\sigma_x) \geq \frac{\sigma_x^2}{K^2\sigma_x^2} = \frac{1}{K^2} \tag{3}$$

This is called **Tchbyshev's Inequality**.

Example 3.1. *Suppose $K = 3$.*

$$P(|X - \mu_x| \geq 3\ \sigma_x) \leq \frac{1}{9}$$

In other words, at least 88% of the observations are within 3 standard deviation from the population mean.

Going back to the our example:

$$X_i \sim (\mu, 1) \quad , \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

We want to study $P(\epsilon \geq \lambda) = P(|\bar{X}_n - \mu| \geq \lambda)$, first we note that:

$$E(X_i) = \mu \quad , \ i = 1, 2, \ldots, n$$

Then:

$$\begin{aligned}
E(\bar{X}_n) &= E\big(\frac{1}{n}\sum_{i=1}^{n} X_i\big) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \frac{1}{\cancel{n}}.(\cancel{n}\mu) \\
&= \mu \qquad (*)
\end{aligned}$$

Thus, using (2) we have:

$$P(|\bar{X}_n - \mu| \geq \delta) \leq \frac{Var(\bar{X}_n)}{\delta^2}$$

Now:

$$\begin{aligned}
Var(\bar{X}_n) &= Var\big(\frac{1}{n}\sum_{i=1}^{n} X_i\big) = \frac{1}{n^2}Var\big(\sum_{i=1}^{n} X_i\big) \\
&= \frac{1}{n^2}\Big[\sum_{i=1}^{n} Var(X_i) + \sum_{1 \leq i < j \leq n}\ \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)\Big] \qquad \textit{using Thm } 5.12(b) - \textit{page } 271 \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \qquad \textit{since } \coprod_{1}^{n} X_i \\
&= \frac{1}{\cancel{n^2}}\cancel{n}Var(X) = \frac{Var(X)}{n} \qquad \textit{since } x_i s \textit{ are identically distributed} \\
&= \frac{\delta_X^2}{n} \qquad (**)
\end{aligned}$$

In our case $X \sim N(\mu, 1)$ so $Var(X) = \delta_X^2 = 1$. Thus $Var(\bar{X}_n) = \frac{1}{n}$

Remark. $X \coprod Y \implies Cov(X,Y) = 0$. *Note that:*

$$X \coprod Y \implies E\big[g_1(X)g_2(Y)\big] = E[g_1(X)].E[g_2(Y)]$$

in particular:

$$X \coprod Y \implies E\big[XY\big] = E[X].E[Y]$$

on the other hand:

$$Cov(X,Y) = E[XY] - E(X)E(Y)$$

thus:

$$X \coprod Y \implies Cov(X,Y) = 0.$$

recall that $X \coprod Y$ means X and Y are independent, i.e. $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ where $f_{X,Y}, f_X and f_Y$ represent respectively the /////TODO
We therefore have:

$$P(|\bar{X}_n - \mu| \geq \delta) \leq \frac{1}{n\delta^2} \tag{4}$$

Using (4) and the sample size, n, we can find an upper bound for the proportion of deviations which are greater than a given threshold δ.
We can also use (4) for Sample Size Deterministic:
Suppose δ is given and we want $P(|\bar{X}_n - \mu| \geq \delta) \leq \beta$ where β is also given. Then setting $\frac{1}{n\delta^2} = \beta$, we can estimate $n \approx \frac{1}{\beta\delta^2}$.

3.3 Application to Voting

Define $X_i = \begin{cases} 1 & \text{NDP} \\ 0 & \text{otherwise} \end{cases}$. Associated to each eligible voter in Canada we have a binary variable X. Let $p = P(X = 1)$. So p represents the proportion of eligible voters who favor NDP. Of interest is often estimation of p. Suppose we have a sample of size n, $X_1, X_2, \ldots, X_n$.
$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample proportion; The counterpart of p which nat be denoted by $\hat{p}$. Note that:

$$\mu_X = E(X) = 1 \times P(X = 1) + 0 \times P(X = 0) = 1 - p + 0 \times (1 - p) = p$$

and:

$$E(X^2) = 1^2 \times P(X = 1) + 0^2 \times P(X = 0) = 1 - p + 0 \times (1 - p) = p$$

From (*) and (**) we find that :

$$E(\hat{p}_n) = E(\bar{X}_n)\mu_X = p$$

and:

$$Var(\hat{p}_n) = E(\bar{X}_n) = \frac{Var(X)}{n} = \frac{\sigma_X^2}{n} = \frac{p(1-p)}{n}$$

Thus using (2), we have:

$$P(|\hat{p}_n - p| \geq \delta) \leq \frac{Var(\hat{p}_n)}{\delta^2} = \frac{p(1-p)}{n\delta^2}$$

Note that the above bound on the probability of derivation depends on p which is *unknown*. We however notice that $p(1-p) \leq \frac{1}{4}$.
Define $\zeta(x) = x(1-x) \ for \ 0 < x < 1$. Then:

$$\zeta^{'}(x) = 1 - 2x \implies \zeta^{'}(x) = 0 \implies x = \frac{1}{2}$$

$$\zeta^{"}(\frac{1}{2}) = -2 \implies x = \frac{1}{2} \qquad \text{which is a } \textbf{maximizer}$$

$$\zeta(\frac{1}{2}) = \frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4}$$

(Note that $\zeta^{"}(x) = -2$ for all $0 < x < 1$)

We therefore find:

$$P(|\hat{p}_n - p| \geq \delta) \leq \frac{1}{4n\delta^2} \tag{5}$$

Using (5) and a given sample size n we can find an upper bound for the probability of derivation by δ and the amount for any given δ.
We can also use (5) for <u>sample size deterministic</u> for a size bound β and derivative δ as follows:

$$\frac{1}{4n\delta^2} = \beta \qquad \implies \qquad n \geq \frac{1}{4\beta\delta^2}$$

This is of course conservative since $p(1-p) \leq \frac{1}{4}$.