Stat 2911 Lecture Notes

Class 10, 2017

Uri Keich

© Uri Keich, The University of Sydney

Markov and Chebyshev's Inequalities, Convergence in probability, Weak Law of Large Numbers, Multinomial distribution (Rice 3.2, 4.1, 4.2, 5.2)

Cor. If X and Y are ind. L^2-RVs, then

$$\mathrm{Cov}(X,Y)=0$$

If $X_1,\dots,X_n \in L^2$ then

$$V\left(\sum_1^n X_i\right)=\sum_1^n V(X_i)+\sum_{i\neq j}\mathrm{Cov}(X_i,X_j)$$

Cor. If X_i are ind. L^2-RVs then

$$V\left(\sum_1^n X_i\right)=\sum_1^n V(X_i)$$

If X_i are iid Bernoulli(p) RVs then

$S_n' = \sum_1^n X_i \sim \mathrm{Binom}(n,p)$ and

$$V(S_n')=np(1-p)$$

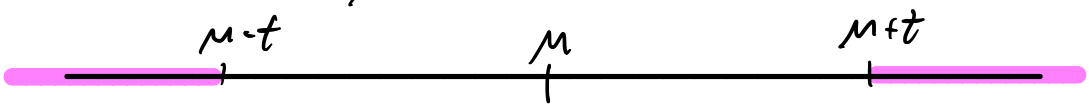
LaTeX workshop Tue 28/3 2-3 pm
Carslaw 610

© Uri Keich, The University of Sydney

Chebyshev's Inequality

Let X be an L^2-RV with $E(X)=\mu$ and $V(X)=\sigma^2$.
For any $t>0$,

$$P(|X-\mu|\geq t)\leq \sigma^2/t^2$$

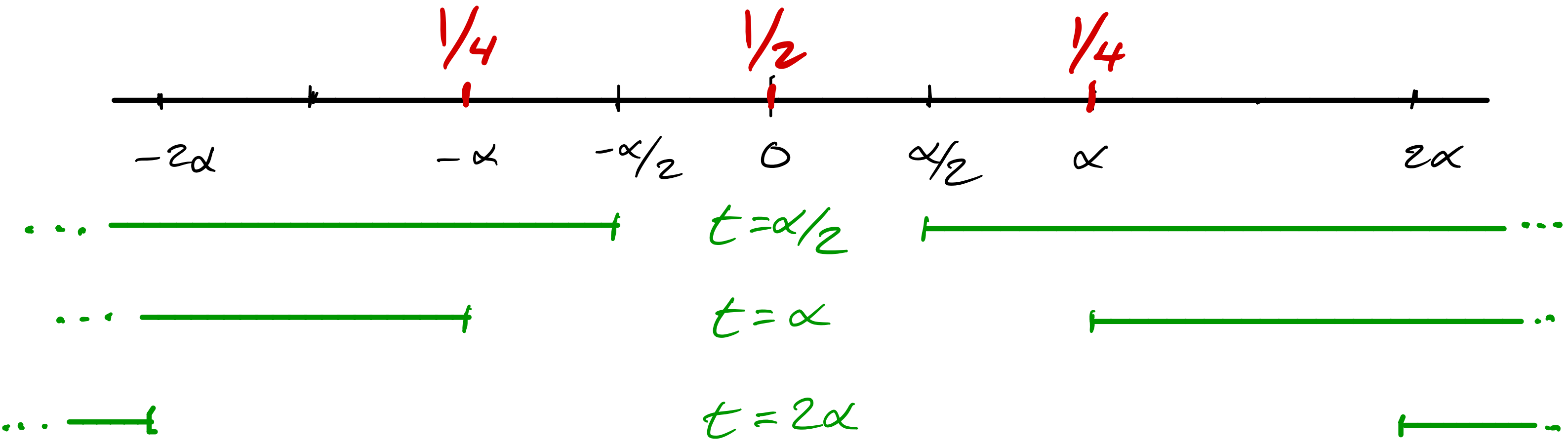


It provides an upper bound on the probability of a deviation from the mean, $(|X-\mu|\geq t)$, requiring very little: only the variance (dispersion about μ)!

Example.

$$p_X(x)=\begin{cases} 1/2 & x=0 \\ 1/4 & x\in\{-\alpha,\alpha\} \\ 0 & x\notin\{-\alpha,0,\alpha\} \end{cases} \qquad \alpha>0$$

$E(X)=$; $E(X^2)=$; $V(X)=$



t	$P(\lvert X-\mu\rvert\geq t)$	σ^2/t^2	Comment
$\alpha/2$	$1/2$	$(\alpha^2/2)/(\alpha/2)^2=2$	useless
α	$1/2$	$(\alpha^2/2)/\alpha^2=1/2$	exact - cannot be improved!
2α	0	$(\alpha^2/2)/(2\alpha)^2=1/8$	informative

Clearly, the utility of Chebyshev's inequality varies.
Maybe we can do better?

Exercise (requires some thought)

Show that σ^2/t^2 is an optimal upper bound.
That is, if $P(|\bar{X}-\mu| \geq t) \leq f(t,\sigma^2)$ for any RV $\bar{X}$ with $V(\bar{X})=\sigma^2$ and any $t>0$, then

$$f(t,\sigma^2) \geq \sigma^2/t^2.$$

Hint: (i) First show that WLOG $\sigma^2=1$ by considering $P(|\bar{X}-\mu| \geq t) = P\left(\left|\frac{\bar{X}-\mu}{\sigma} - 0\right| \geq t/\sigma\right)$

(ii) Use a similar example to the exact one above

<u>Lemma</u>. Markov's Inequality

Let $\underline{X}$ be a non-negative RV $(P(\underline{X} \geq 0) = 1)$, then $\forall t > 0$

$$P(\underline{X} \geq t) \leq E(\underline{X})/t$$

<u>Proof</u>.

$\underline{X} \geq 0$ so we can go ahead and compute

$$E(\underline{X}) = \sum_k x_k \, p_X(x_k)$$

$$\geq \sum_{k:\, x_k \geq t} x_k \, p_X(x_k)$$

$$\geq \sum_{k:\, x_k \geq t} t \, p_X(x_k)$$

$$= t \, P(\underline{X} \geq t) \qquad \square$$

Where have we used the fact that $P(\underline{X} \geq 0) = 1$?

<u>Proof</u> (of Chebyshev's Inequality).

$$P(|\underline{X} - \mu| \geq t) = P[(\underline{X} - \mu)^2 \geq t^2]$$

$$\stackrel{?}{\leq} E(\underline{X} - \mu)^2 / t^2$$

$$= \sigma^2 / t^2$$

The law of large numbers (a special case)

Let $X_1, X_2, \dots$ be iid Bernoulli(p) RVs and let

$S_n = \sum_1^n X_i$, then $S_n \sim$

$$E(S_n/n) = \tfrac{1}{n} E(S_n) =$$

$$V(S_n/n) = \tfrac{1}{n^2} V(S_n) = \qquad \xrightarrow[n]{}$$

Intuitively, $S_n/n \longrightarrow p = E(S_n/n) = E(X_i)$

Def. A sequence of RVs Y_n converges to the RV Y in probability, if $\forall \varepsilon > 0$,

$$P(|Y_n - Y| > \varepsilon) \xrightarrow[n]{} 0$$

Claim. S_n/n converges to $E(X_i) = p$ in probability.

A special case of the weak law of large numbers (WLLN).

Proof. $\forall \varepsilon > 0$,

$$P(|S_n/n - p| \geq \varepsilon) \overset{?}{\leq} \frac{V(S_n/n)}{\varepsilon^2}$$

$$= \frac{\frac{1}{n} p(1-p)}{\varepsilon^2} \xrightarrow[n]{}$$

© Uri Keich, The University of Sydney

<u>Theorem</u> (WLLN):

If $X_1, X_2, \ldots$ are iid L^2-RVs with mean μ, and variance σ^2. Then with $S_n = \sum_1^n X_i$,

$$S_n/n \longrightarrow \mu \text{ in probability.}$$

Is it really necessary for $X_i \in L^2$?

The strong law of large numbers (SLLN) states that if X_i are L^1-RVs then

$$P(S_n/n \longrightarrow \mu) = 1$$

<u>The multinomial distribution</u> models an experiment with n iid trials. Each trial has r possible outcomes, where outcome i has probability p_i so that $\sum_1^r p_i = 1$.

<u>Example</u>. Roll a die (fair or not) n times and record the n outcomes.

The <u>multinomial random vector</u> $\underset{\sim}{N} = (N_1, \dots, N_r)$ yields the number of trials each outcome came up.

<u>Example</u>. For the sequence of rolls

$$3, 6, 2, 2, 1, 6, 6, 3, 4, 1, 1, 5, 4, 3, 6$$

$$\underset{\sim}{N} = (3, 2, 3, 2, 1, 4)$$

Note that we already encountered the multinomial distribution: the case $r=2$ is the binomial dist.

Goal: Find the pmf $p_{\underset{\sim}{N}}(n_1, \dots, n_r) = P(N_1 = n_1, \dots, N_r = n_r)$.

Clearly, $n_i \in \mathbb{N}$ and $\sum_1^r n_i = n$ or else $p_{\underset{\sim}{N}}(\underset{\sim}{n}) = 0$.

$\{N_1 = n_1, \dots, N_r = n_r\} \iff$ outcome/type i appeared n_i times / in n_i trials

© Uri Keich, The University of Sydney

One such configurations of trials is given by

$$\underbrace{1 1 \dots 1}_{n_1} \; \underbrace{2 \dots 2}_{n_2} \; \dots \; \underbrace{r \dots r}_{n_r}$$

The prob. of this particular configuration is

$$p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

What is the probability of another config. that differs only by the order of the outcomes?

How many different such configurations are there?

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-\dots-n_{r-1}}{n_r}$$

$$\overset{\text{ex}}{=} \frac{n!}{n_1! \dots n_r!} \overset{\text{notation}}{=:} \underbrace{\binom{n}{n_1\, n_2 \, \dots \, n_r}}_{\text{the multinomial coefficient}}$$

$$\Rightarrow p_{\underline{N}}(n_1, \dots, n_r) = \binom{n}{n_1 \dots n_r} p_1^{n_1} \cdots p_r^{n_r}$$

Sum the probs of all configs with n_i outcomes of type i

Does it agree with the binomial pmf?

Note: $\binom{n}{n_1 \dots n_r}$ = # of ways to order n balls in a row where n_i balls are of color i (= the type i)

Alternatively, using

$$P(A_1 \cap \dots \cap A_r) \overset{ex}{=} P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_r|A_1 \cap \dots \cap A_{r-1})$$

$$\begin{aligned} P(N_1 = n_1, \dots, N_r = n_r) = {} & P(N_1 = n_1) \\ & \cdot P(N_2 = n_2 | N_1 = n_1) \\ & \cdot P(N_3 = n_3 | N_1 = n_1, N_2 = n_2) \\ & \dots \\ & \cdot P(N_r = n_r | N_1 = n_1, N_2 = n_2, \dots, N_{r-1} = n_{r-1}) \end{aligned}$$

What is the dist. of N_1?

$$N_1 \sim \text{binom}(n, p_1)$$

What is the dist. of N_2 given that $N_1 = n_1$?

$$P(N_2 = n_2 | N_1 = n_1) = \frac{P(N_1 = n_1, N_2 = n_2)}{P(N_1 = n_1)} = \frac{\binom{n}{n_1\ n_2} p_1^{n_1} p_2^{n_2} (1 - p_1 - p_2)^{n - n_1 - n_2}}{\binom{n}{n_1} p_1^{n_1} (1 - p_1)^{n - n_1}}$$

$$\overset{\text{exercise}}{=} \dots = \binom{n - n_1}{n_2} \left(\frac{p_2}{1 - p_1} \right)^{n_2} \left(1 - \frac{p_2}{1 - p_1} \right)^{(n - n_1) - n_2}$$

$$\Rightarrow \quad N_2 | N_1 = n_1 \sim \text{binom}(n - n_1, p_2/(1 - p_1))$$

which in hindsight is intuitively obvious

What is the dist. of N_3 given that $N_1 = n_1$ and $N_2 = n_2$?

$$N_3 | N_1 = n_1, N_2 = n_2 \sim \text{binom}(n - n_1 - n_2, p_3/(1 - p_1 - p_2))$$

$\vdots$

$$P(N_1=n_1,\ \dots,\ N_r=n_r)=\binom{n}{n_1}p_1^{n_1}(1-p_1)^{n-n_1}$$

$$\cdot\binom{n-n_1}{n_2}\left(\frac{p_2}{1-p_1}\right)^{n_2}\left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-n_1-n_2}$$

$$\cdot\binom{n-n_1-n_2}{n_3}\left(\frac{p_3}{1-p_1-p_2}\right)^{n_3}\left(\frac{1-p_1-p_2-p_3}{1-p_1-p_2}\right)^{n-n_1-n_2-n_3}$$

$$\vdots$$

$$\cdot\binom{n-n_1-n_2-\dots-n_{r-1}}{n_r}\underbrace{\left(\frac{p_r}{1-p_1-\dots-p_{r-1}}\right)^{n_r}}_{p_r/p_r=1}\underbrace{\left(\frac{1-p_1-\dots-p_r}{1-p_1-\dots-p_{r-1}}\right)^{n-n_1-\dots-n_r}}_{0^0=1 \text{ here}}$$

exercise

$$=\binom{n}{n_1\ \dots\ n_r}p_1^{n_1}\cdots p_r^{n_r}.$$

© Uri Keich, The University of Sydney