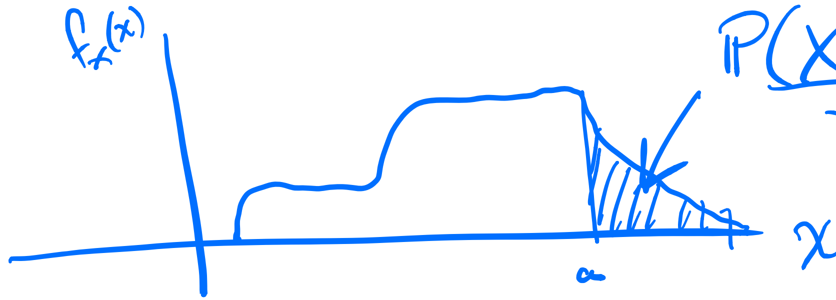
Markov, Chebyshev, and the Law of Large Numbers

Lec.24

August 3, 2020

Markov's Inequality: Fundamental Idea



$\mathbb{P}(X \geq a) \leq ?$

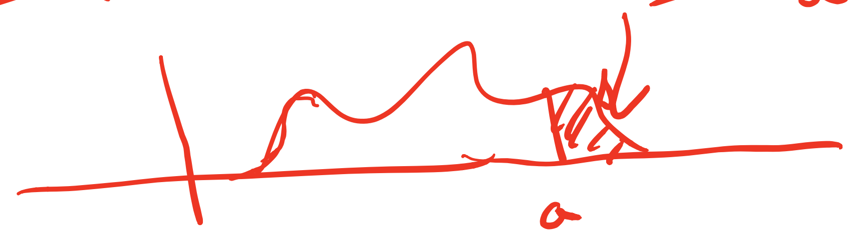
Simple bound on the tail of a random variable, that uses only the expected value (first moment), and the fact that the random variable is nonnnegative.

Markov's Inequality: Definition

If X is a nonnegative random variable with finite mean and $a > 0$, then the probability that X is at least a is at most the expectation of X divided by a.

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

$$\leq \frac{\mathbb{E}[X]}{a}$$



Markov's Inequality: Proof I

WLOG, let X be a nonnegative continuous R.V.

$f(x) = f_X(x)$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_0^{\infty} x \cdot f(x)\,dx$$

$$= \underbrace{\int_0^{a} x \cdot f(x)\,dx}_{\text{nonnegative}} + \int_a^{\infty} x \cdot f(x)\,dx$$

$$\geq \int_a^{\infty} x \cdot f(x)\,dx$$

x is at least as big as a

$$\geq \int_a^{\infty} a \cdot f(x)\,dx$$

$$= a \cdot \underbrace{\int_a^{\infty} f(x)\,dx}$$

$$= a \cdot \mathbb{P}(X \geq a)$$

$$\Rightarrow \mathbb{E}[X] \geq a \cdot \mathbb{P}(X \geq a) \Rightarrow \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

"If $\mathbb{E}[X]$ is small, then the prob. that X is large, is small"*

* assuming X is nonnegative.

Markov's Inequality: Proof I

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} xf(x)dx && (1)\\
&= \int_{0}^{a} xf(x)dx + \int_{a}^{\infty} xf(x)dx && (2)\\
&\geq \int_{a}^{\infty} xf(x)dx && (3)\\
&\geq \int_{a}^{\infty} af(x)dx && (4)\\
&= a \cdot \int_{a}^{\infty} f(x)dx && (5)\\
&= aP(X \geq a) && (6)\\
& && (7)
\end{aligned}
$$

Thus, $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$

Markov's Inequality: Proof II

Let I be the indicator r.v. defined as follows:

$$I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{o.w.} \end{cases}$$

Then

$$X \geq a \cdot I$$

only holds if X is nonnegative. To prove this, look at the cases $X \geq a$, $X < a$

$$\mathbb{E}[X] \geq \mathbb{E}[a \cdot I]$$

$$\mathbb{E}[X] \geq a \cdot \mathbb{E}[I]$$

$$\mathbb{E}[X] \geq a \cdot \underline{P(X \geq a)}$$

$$\Rightarrow \quad P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Markov's Inequality: Proof II

Let I be the indicator r.v. defined as follows:

$$I = \begin{cases} 1, \text{if } X \geq a \\ 0, \text{o.w.} \end{cases} \tag{8}$$

Then,

$$\begin{aligned} X &\geq a \cdot I & (9) \\ \mathbb{E}[X] &\geq \mathbb{E}[a \cdot I] & (10) \\ \mathbb{E}[X] &\geq a\mathbb{E}[I] & (11) \\ \mathbb{E}[X] &\geq aP(X \geq a) & (12) \end{aligned}$$

Thus, $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$

Markov's Inequality: Proof III

$$\mathbb{E}[X] = \underbrace{\mathbb{E}[X \mid X<a] \cdot \mathbb{P}(\overbrace{X<a}^{A})}_{\text{nonnegative}} + \mathbb{E}[X \mid X \geq a] \cdot \mathbb{P}(\overbrace{X \geq a}^{A^c})$$

$$\geq \underbrace{\mathbb{E}[X \mid X \geq a]} \cdot \mathbb{P}(X \geq a)$$

$$\geq a \cdot \mathbb{P}(X \geq a)$$

$$\implies \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Markov's Inequality: Proof III

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X|X < a] \cdot P(X < a) + \mathbb{E}[X|X \geq a] \cdot P(X \geq a) && (13)\\
&\geq \mathbb{E}[X|X \geq a] \cdot P(X \geq a) && (14)\\
&\geq a \cdot P(X \geq a) && (15)
\end{aligned}$$

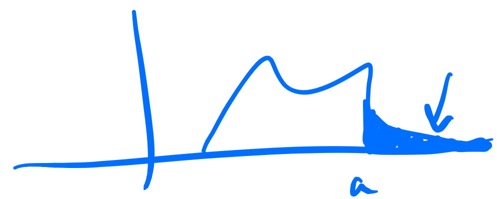
Thus, $P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$

Example: Markov & Coin Flips

Let $X \sim Geom(\frac{1}{2})$. Use Markov's inequality to upper bound $P(X > 10)$.

$$\mathbb{P}(X>10) = \mathbb{P}(X \geq 11) \leq \frac{\mathbb{E}[X]}{11} = \frac{2}{11}$$

What is $\mathbb{P}(X \geq 11)$ exactly?



$$\mathbb{P}(X \geq 11) = \left(\frac{1}{2}\right)^{10}$$

$$= \frac{1}{2^{10}}.$$

$$\frac{1}{2^{10}} << \frac{2}{11},$$

so Markov's inequality gives a <u>loose</u> bound in this case.

Example: Markov & Coin Flips

Let $X \sim Geom(\frac{1}{2})$. Use Markov's inequality to upper bound $P(X > 10)$.

$P(X \geq 11)$

$$P(X > 10) \leq \frac{\mathbb{E}[X]}{11} = \frac{2}{11} \qquad (16)$$

If we try to actually calculate $P(X > 10)$:

$$P(X > 10) = (1 - p)^{10} = (\frac{1}{2})^{10} \qquad (17)$$

Note that $\frac{1}{2^{10}} << \frac{2}{11}$, so Markov's bound can be pretty loose.

Generalized Markov's Inequality: Definition

If X is **any** random variable with finite mean and $a > 0$, then for any $r > 0$:

$$P(|X| \geq a) \leq \frac{\mathbb{E}[|X|^r]}{a^r}$$

Proof: Try it yourself, then see notes.

Chebyshev's Inequality: Fundamental Idea

Often times we can do better than Markov's Inequality if we use more information about the random variable. For this inequality, we use the first two moments, $E[X]$ and $E[X^2]$.

Note: The variance of a random variable captures these two moments, and is related to how much probability there is in the tails.

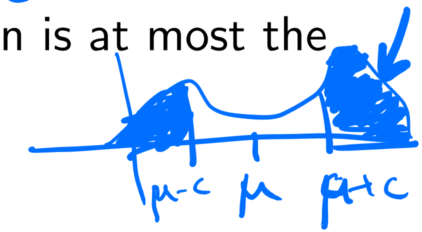
$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Chebyshev's Inequality: Definition

doesn't need to be nonnegative.

and finite variance.

If X is a random variable with finite mean μ and $c > 0$, then the probability that X is at least c away from its mean is at most the variance of X divided by c^2.



$c \in \mathbb{R}$

$c > 0$

$$P(|X - \mu| \geq c) \leq \frac{\mathrm{Var}[X]}{c^2}$$

upper bound on both tails.

Note: X does not need to be nonnegative in order to apply Chebyshev's inequality. c is a positive constant.

Chebyshev's Inequality: Proof

$\mathbb{E}[X] = \mu$

Let $Y = (X-\mu)^2$. then Y is always ≥ 0

$\mathbb{E}[Y] = \mathbb{E}[(X - \mathbb{E}[X])^2] = Var[X]$

Note that $\mathbb{P}(|X-\mu| \geq c) = \mathbb{P}(Y \geq c^2)$

Markov's

So,

$$\mathbb{P}(|X-\mu| \geq c) = \mathbb{P}(Y \geq c^2) \leq \frac{\mathbb{E}[Y]}{c^2} = \frac{Var[X]}{c^2}$$

$$\Rightarrow \mathbb{P}(|X-\mu| \geq c) \leq \frac{Var[X]}{c^2}$$

"If the variance of X is small then the probability X is far from its mean is small"

Chebyshev's Inequality: Proof

Define $Y = (X - \mu)^2$ and note that $\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \mathrm{Var}[X]$. Also, notice that the event that we are interested in, $|X - \mu| \geq c$, is exactly the same as the event $Y = (X - \mu)^2 \geq c^2$. Therefore, $\Pr[|X - \mu| \geq c] = \Pr[Y \geq c^2]$.

Moreover, Y is always nonnegative, so we can apply Markov's inequality to get

$$\Pr[|X - \mu| \geq c] = \Pr[Y \geq c^2] \leq \frac{\mathbb{E}[Y]}{c^2} = \frac{\mathrm{Var}[X]}{c^2}.$$

Example: Chebyshev & Coin Flips

$\mathbb{P}(X > 10)$ vs $\mathbb{P}(X \geq 10)$

Let $X \sim Geom(\frac{1}{2})$. Use Chebyshev's inequality to upper bound $P(X > 10)$.

$E[X] = 2 = \mu$

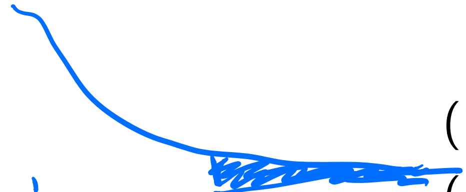
$Var[X] = 2$

$$\underline{\mathbb{P}(X > 10)} = \mathbb{P}(X \geq 11)$$
$$= \mathbb{P}(X \geq \mu + 9)$$
$$= \mathbb{P}(\underline{X - \mu \geq 9})$$
$$\leq \mathbb{P}(|X - \mu| \geq 9)$$
$$\leq \frac{Var(X)}{9^2} = \frac{2}{81}$$

This is tighter than Markov's $\left(\frac{2}{11}\right)$ but it is still far off from $\frac{1}{2^{10}}$

Example: Chebyshev & Coin Flips

Let $X \sim Geom(\frac{1}{2})$. Use Chebyshev's inequality to upper bound $P(X > 10)$.



$$\mathbb{E}[X] = \mu = 2 \tag{18}$$

$$\mathrm{Var}[X] = 2 \tag{19}$$

$$P(X > 10) = P(X > \mu + 8) = P(X - \mu > 8) \tag{20}$$

$$P(X > 10) \leq P(|X - \mu| > 8) = P(|X - \mu| \geq 9) \tag{21}$$

$$\leq \frac{\mathrm{Var}[X]}{\cancel{8}\,9^2} = \frac{2}{\cancel{64}\,81} \cancel{= \frac{1}{32}} \tag{22}$$

This is a tighter bound than Markov's $(\frac{1}{5})$, but is still far off from the true probability $\dfrac{1}{2^{10}}$.

Chebyshev Corollary

For any random variable X with finite expectation $\mathbb{E}[X] = \mu$ and finite standard deviation $\sigma = \sqrt{\mathrm{Var}[X]}$,

$$\underbrace{\Pr[\,|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}}\,,$$

for any constant $k > 0$.

$$\frac{Var(X)}{c^2} = \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$$

Proof:

Plug $= k\sigma$ into Chebyshev's inequality.

$c = k\sigma$

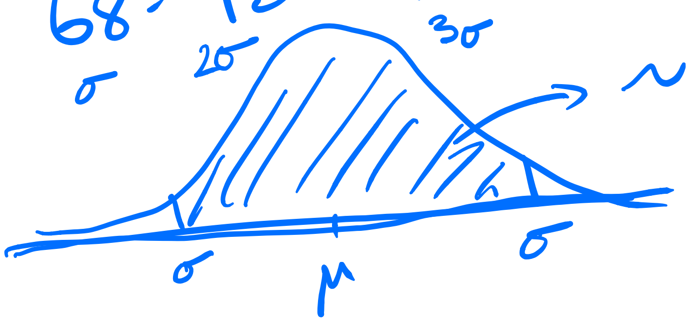
Chebyshev Corollary: Example

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Find a bound on the probability that X is 2σ or more away from its mean μ.

$$P(|X - \mu| \geq \underset{k}{2}\sigma) \leq \frac{1}{2^2} = \frac{1}{4}$$

Note:

68-95-99.7 rule.



95% w/ in 2 std dev.

5% prob of being outside 2 SD $= \frac{1}{20} \ll \frac{1}{4}$.

Chebyshev Corollary: Example

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Find a bound on the probability that X is 2σ or more away from its mean μ.

$$\Pr[\,|X - \mu| \geq 2\sigma] \leq \frac{1}{2^2} = \frac{1}{4}$$

Note: Our empirical 68–95–99.7 rule for normal distributions indicates that this can be quite a crude bound. This empirical rule says 95% of the time X will fall within two standard deviations, meaning it will fall 2σ away from its mean μ with probability 5%.

Law of Large Numbers: Fundamental Idea

observe random variables ⟹ data.

If we observe a random variable many times, and average our observations, then the average will converge to the average of the random variable.

Law of Large Numbers: Definition

Let $X_1, X_2, \ldots,$ be a sequence of i.i.d. random variables with common finite expectation $\mathbb{E}[X_i] = \mu$ and variance $\mathrm{Var}[X_i] = \sigma^2$ for all i. Then, their partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfy

$$\Pr\left[\left|\frac{1}{n}S_n - \mu\right| < \epsilon\right] \to 1 \qquad \text{as } n \to \infty,$$

for every $\epsilon > 0$, however small.

$\frac{1}{n} S_n$ "sample mean"

"sample mean converges to the mean"

Law of Large Numbers: Proof

$\varepsilon > 0$

$$\mathbb{P}\left(\left| \underbrace{\tfrac{1}{n} S_n}_{} - \mu \right| \geq \varepsilon \right) \leq \frac{\mathrm{Var}(\frac{1}{n}S_n)}{\varepsilon^2} = \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

goes to 0 as $n \to \infty$

$\frac{1}{n} S_n$ is a random variable

$$\mathbb{E}\left[\tfrac{1}{n} S_n\right] = \frac{1}{n} \sum \mathbb{E}[X_i] = \frac{n \cdot \mu}{n} = \mu$$

$$\mathrm{Var}\left[\tfrac{1}{n} S_n\right] = \frac{1}{n^2} \cdot \sum \mathrm{Var}[X_i] = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

$$\Rightarrow \mathbb{P}\left(\left| \tfrac{1}{n} S_n - \mu \right| < \varepsilon \right) = 1 - \overbrace{\mathbb{P}\left(\left| \tfrac{1}{n} S_n - \mu \right| \geq \varepsilon \right)}^{\to 0 \text{ as } n \to \infty}$$

$\to 1$

as $n \to \infty$

Law of Large Numbers: Proof

Let $\mathrm{Var}[X_i] = \sigma^2 < \infty$ be the common variance of the r.v.'s. Since $X_1, X_2, \ldots$ are i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$, we have $\mathbb{E}[\frac{1}{n}S_n] = \mu$ and $\mathrm{Var}[\frac{1}{n}S_n] = \frac{\sigma^2}{n}$, so by Chebyshev's inequality we have

$$\Pr\left[|\tfrac{1}{n}S_n - \mu| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\frac{1}{n}S_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \qquad \text{as } n \to \infty.$$

Hence, $\Pr\left[|\frac{1}{n}S_n - \mu| < \epsilon\right] = 1 - \Pr\left[|\frac{1}{n}S_n - \mu| \geq \epsilon\right] \to 1$ as $n \to \infty$.

Example: Law of Large Numbers

$\mathbb{E}[X_i] = \frac{1}{2}$

$X_i \sim \text{Bern}\left(\frac{1}{2}\right)$

Sample mean

flips: $0,\ 1,\ 1,\ 0,\ 1 \rightarrow \underline{\frac{1}{n} \cdot S_n} = \frac{3}{5}$

Consider a series of coin flips, where each coin flip is independent of the others and has distribution *Bernoulli*(1/2), where 1 corresponds to heads and 0 corresponds to tails.

Example: Law of Large Numbers

Consider a series of coin flips, where each coin flip is independent of the others and has distribution *Bernoulli*$(1/2)$, where 1 corresponds to heads and 0 corresponds to tails.

The Law of Large Numbers states that the proportion of heads is likely to be near $1/2$, the true mean, for a large number of flips.