MAT330/681 Lecture 21 4/21/2021

Q: What can be said about random variables when you don't know their distribution?

Markov Inequality

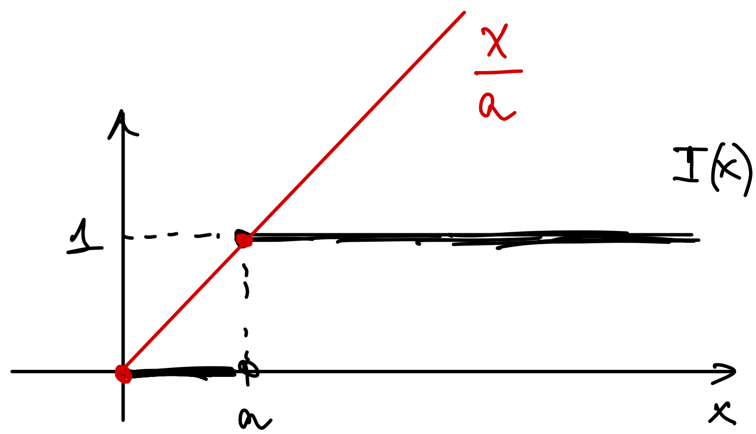
Prop: If X is a random variable that only assumes nonnegative values, then for all $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Pf: Let $I : [0,+\infty) \to \{0,1\}$ be the indicator function of the half-line $[a,+\infty)$:

$$I : [0,+\infty) \longrightarrow \{0,1\}$$

$$I(x) = \begin{cases} 1 & \text{if} \quad x \geq a \\ 0 & \text{otherwise} \end{cases}$$



$$E(I(X)) = \int_0^{+\infty} \underbrace{I(x)}_{\substack{\equiv 0 \text{ if } x<a \\ \equiv 1 \text{ if } x \geq a}} f(x)\,dx = \int_a^{+\infty} f(x)\,dx = P(X \geq a)$$

Since $I(x) \leq \frac{x}{a}$ for all $x \geq 0$, taking expected values:

$$P(X \geq a) = E(I(X)) \leq E\left(\frac{X}{a}\right) \overset{\text{linearity}}{=} \frac{1}{a} E(X).$$

$\square$

Chebyshev's Inequality.

Prop. If X is a random variable with finite mean $\mu = E(X)$, and variance $\sigma^2 = Var(X)$, then $\forall k > 0$:

$$P(\underbrace{|X-\mu| \geq k}) \leq \frac{\sigma^2}{k^2}.$$

Measure of how much X deviates from its mean

Recall:
$Var(X) = E((X-\mu)^2)$

Pf: $|X-\mu| \geq k \iff |X-\mu|^2 \geq k^2 = a$

Apply Markov's Inequality with $a = k^2$:

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

$\square$

Ex: Suppose that a fishing boat collects 50 fish each week, on average. There is no information on the exact distribution.

a) How large is the probability that the fishing boat collects more than 75 fish in a given week?

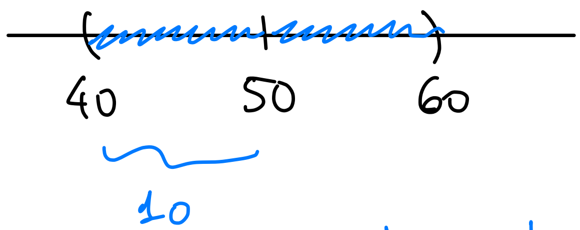
X = # fish collected on a given week. (≥ 0)

$$P(X \geq 75) \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}.$$

b) If the variance in the number of fish collected each week is 25, find a lower bound for the probability that the number of fish collected in a week is between 40 and 60.

$\mu = E(X) = 50, \quad \sigma^2 = Var(X) = 25.$

$$P(40 \le X \le 60) = P(|X-\mu| \le 10) = 1 - P(|X-\mu| \ge 10) \ge 1 - \frac{1}{4} = \frac{3}{4}.$$



b/c by Chebyshev's ineq:

$$P(|X-\mu| \ge 10) \le \frac{\sigma^2}{100} = \frac{25}{100} = \frac{1}{4}.$$

$40 \le x \le 60 \iff |x-50| \le 10$

Weak Law of Large Numbers:

Prop: Let $X_1, X_2, X_3, \dots$ be a sequence of independent identically distributed (iid) random variables with finite mean $\mu = E(X_i)$. Then $\forall \varepsilon > 0$

$$\lim_{n \to \infty} P\left(\left| \underbrace{\frac{X_1 + X_2 + X_3 + \dots + X_n}{n}}_{\overline{X}_n} - \mu \right| \ge \varepsilon \right) = 0.$$

Pf: (Assume that $\sigma^2 < \infty$). Let $\overline{X}_n := \dfrac{X_1 + X_2 + \dots + X_n}{n}$

$$E(\overline{X_n}) = E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) \overset{\substack{\text{linearity}\\ \downarrow}}{=} \frac{E(X_1)+E(X_2)+\cdots+E(X_n)}{n} \overset{\substack{\text{iid}\\ \downarrow}}{=} \frac{\cancel{n}\cdot\mu}{\cancel{n}}$$

$$Var(\overline{X_n}) = Var\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) \overset{\substack{\text{indep}\\ \downarrow}}{=} \frac{Var(X_1)+Var(X_2)+\cdots+Var(X_n)}{n^2}$$

$$\overset{\substack{\text{iid}\\ \downarrow}}{=} \frac{n\cdot\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Using Chebyshev's inequality on $\overline{X_n}$:

$$0 \leq P(|\overline{X_n}-\mu| \geq \varepsilon) = P(|\overline{X_n} - \underbrace{E(\overline{X_n})}_{=\mu}| \geq \varepsilon) \leq \frac{\overbrace{Var(\overline{X_n})}^{=\sigma^2/n}}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

By Squeeze Thm.

$$0 \leq \lim_{n\to\infty} P(|\overline{X_n}-\mu| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0$$

so $\lim_{n\to\infty} P(|\overline{X_n}-\mu| \geq \varepsilon) = 0$ as desired. $\square$

Rmk: The above means that $\overline{X_n}$ converges in probability to μ.
$\left(\overline{X_n} \xrightarrow{P} \mu\right)$.

(Stay tuned for the Strong Law of Large Numbers in the Next Lecture)

Central Limit Theorem (Baby version).

Let $X_1, X_2, X_3, \ldots$ be a sequence of iid random variables with mean $\mu = E(X_i)$ and variance $\sigma^2 = Var(X_i)$.

Then for all $a \in \mathbb{R}$

standard normal random variable $Z \sim Normal(0,1)$.

$$\lim_{n\to\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq a\right) = P(Z \leq a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-z^2/2}\, dz.$$

Lemma: Let $Z_1, Z_2, \ldots$ be a sequence of random variables with c.d.f. F_{Z_n} and moment generating function M_{Z_n}. Let Z be a random variable with c.d.f. F_Z and moment gen. function M_Z. If $M_{Z_n}(t) \to M_Z(t)$ for all t, then $F_{Z_n}(t) \to F_Z(t)$ for all t at which F_Z is continuous.

From last class: If $Z \sim Normal(0,1)$, then $M_Z(t) = e^{t^2/2}$.

We will use the above Lemma as follows:

If $Z_1, Z_2, \ldots, Z_n, \ldots$ is a sequence of random variables s.t.

$M_{Z_n}(t) \longrightarrow e^{t^2/2}$ for all t, then

$$F_{Z_n}(t) \longrightarrow F_Z(t) = P(Z \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-z^2/2}\, dz.$$

Pf of CLT: Assume $\mu = 0$ and $\sigma = 1$.

Let $M(t)$ be the moment generating function of X_i.

$M(0) = 1$.

$M'(0) = E(X_i) = \mu = 0$

$M''(0) = E(X_i^2) = 1$

$$\left(\begin{array}{l} Var(X_i) = \sigma^2 = 1 \\ \quad \| \\ E(X_i^2) - \underbrace{E(X_i)^2}_{0} \end{array} \right)$$

Moreover, the moment generating function of $\frac{X_i}{\sqrt{n}}$ is

$$E\left(e^{t\frac{X_i}{\sqrt{n}}}\right) = E\left(e^{\frac{t}{\sqrt{n}}X_i}\right) = M\left(\frac{t}{\sqrt{n}}\right). \quad (*)$$

By independence of X_i; the moment generating function of

$$\frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}} = \frac{X_1}{\sqrt{n}} + \frac{X_2}{\sqrt{n}} + \cdots + \frac{X_n}{\sqrt{n}} \quad \text{is}$$

$$M_{\frac{X_1}{\sqrt{n}}}(t) \cdot M_{\frac{X_2}{\sqrt{n}}}(t) \cdot \ldots M_{\frac{X_n}{\sqrt{n}}}(t) \overset{(*)}{=} \underbrace{M\left(\frac{t}{\sqrt{n}}\right) \cdot M\left(\frac{t}{\sqrt{n}}\right) \cdots M\left(\frac{t}{\sqrt{n}}\right)}_{n}$$

$$= \left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Note:

$$\frac{\overline{X_n} - \overset{0}{\cancel{\mu}}}{\underset{1}{\cancel{\sigma}}/\sqrt{n}} = \frac{\overline{X_n}}{1/\sqrt{n}} = \frac{\frac{X_1 + X_2 + \cdots + X_n}{n}}{1/\sqrt{n}} = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

By the Lemma, it suffices to show

$$\lim_{n\to\infty}\left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n = e^{t^2/2} = M_Z(t)$$

Let $L(t) := \log M(t)$.

Then $L(0) = \log M(0) = 0$.

Recall:
$M(0) = 1$.
$M'(0) = E(X_i) = \mu = 0$
$M''(0) = E(X_i^2) = 1$

$$L'(0) = \frac{1}{M(t)}\cdot M'(t)\Big|_{t=0} = \frac{M'(0)}{M(0)} = \frac{0}{1} = 0.$$

$$L''(0) = \frac{d}{dt}\left(\frac{M'(t)}{M(t)}\right)\Big|_{t=0} = \frac{\cancelto{1}{M''(0)}\cdot \cancelto{1}{M(0)} - \left(\cancelto{0}{M'(0)}\right)^2}{\cancelto{1}{M(0)}^2} = 1.$$

Now,

$$\lim_{n\to\infty}\frac{L(t/\sqrt{n})}{n^{-1}} \overset{\text{L'Hospital}}{=} \lim_{n\to\infty}\frac{L'(t/\sqrt{n})\cdot t\cdot \cancel{n^{-3/2}}\cdot \cancel{n^{3/2}}}{2n^{-2}\cdot n^{3/2}}$$

$$= \lim_{n\to\infty}\frac{L'(t/\sqrt{n})\cdot t}{2n^{-1/2}} \overset{\text{L'Hospital}}{=} \lim_{n\to\infty}\frac{\overbrace{L''(t/\sqrt{n})}^{L''(0)=1}\cdot t^2\,\cancel{n^{-3/2}}}{2\cancel{n^{-3/2}}}$$

$$= \frac{t^2}{2}; \quad \text{for all } t.$$

So, for all t, we have:

$$\left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(e^{L(t/\sqrt{n})}\right)^n = e^{nL(t/\sqrt{n})} \xrightarrow{n\to\infty} e^{t^2/2} = M_Z(t).$$

The general version (without assuming $\mu = 0$ and $\sigma = 1$) follows from applying the above version to the "standardized" random variables $\tilde{X}_i := \dfrac{X_i - \mu}{\sigma}$, noting that

$E(\tilde{X}_i) = 0$ and $Var(\tilde{X}_i) = 1$.

$\square$

<u>Rmk</u> (If you know some Real Analysis): The convergence in the CLT:

$$\lim_{n \to \infty} P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq a\right) = P(Z \leq a)$$

is not only <u>pointwise</u> but also <u>uniform</u> in a.