$$\sum(Z_i - \bar{Z})^2 = \sum Z_i^2 - 2\sum Z_i \bar{Z} + \sum \bar{Z}^2$$
$$= \sum Z_i^2 - 2n\bar{Z}^2 + n\bar{Z}^2 = \sum Z_i^2 - n\bar{Z}^2$$
$$= \vec{Z}^T \vec{Z} - \vec{Z}^T\left(\tfrac{1}{n}J_n\right)\vec{Z}$$
$$= \vec{Z}^T I \vec{Z} - \vec{Z}^T\left(\tfrac{1}{n}J_n\right)\vec{Z}$$
$$= \vec{Z}^T \underbrace{\left(I - \tfrac{1}{n}J_n\right)}_{B_1} \vec{Z}$$

Can we use Cochran's thm here? We need to verify the two conditions

$$I \stackrel{?}{=} B_1 + B_2 = \left(I - \cancel{\tfrac{1}{n}J_n}\right) + \cancel{\tfrac{1}{n}J_n} = I \quad \text{yes!}$$
$$n \stackrel{?}{=} \text{rank}[B_1] + \text{rank}[B_2] = (n-1) + 1 = n \quad \text{yes!}$$

Thm: if matrix A is symmetric and idempotent i.e. AA = A then rank[A] = tr[A] := sum of A's diagonal entries

Let's verify both B_1 and B_2 are idempotent

$$B_2 B_2 = \left(\tfrac{1}{n}J_n\right)\left(\tfrac{1}{n}J_n\right) = \tfrac{1}{n^2}J_n J_n = \tfrac{1}{n^2}\begin{bmatrix} n & n & \cdots & n \\ n & n & \cdots & \\ \vdots & & \ddots & \\ n & & & n \end{bmatrix} = \tfrac{1}{n}J = B_2 \checkmark$$
$$B_1 B_1 = \left(I - \tfrac{1}{n}J_n\right)\left(I - \tfrac{1}{n}J_n\right) = II - \tfrac{1}{n}J_n I - I\tfrac{1}{n}J_n + \tfrac{1}{n}J_n \tfrac{1}{n}J_n$$
$$= I - \tfrac{1}{n}J_n - \cancel{\tfrac{1}{n}J_n} + \cancel{\tfrac{1}{n}J_n} = B_1 \checkmark$$

Verify both B_1 and B_2 are symmetric

$$B_2^T = \left(\tfrac{1}{n}J_n\right)^T = \tfrac{1}{n}J_n^T = \tfrac{1}{n}J_n = B_2 \checkmark$$
$$B_1^T = \left(I - \tfrac{1}{n}J_n\right)^T = I^T - \left(\tfrac{1}{n}J_n\right)^T = I - \tfrac{1}{n}J_n = B_1 \checkmark$$
$$\Rightarrow \text{rank}[B_2] = \text{tr}[B_2] = \sum_{i=1}^n B_{2_{i,i}} = \tfrac{1}{n}\sum_{i=1}^n J_{n_{i,i}} = \tfrac{1}{n}n = 1$$
$$\Rightarrow \text{rank}[B_1] = \text{tr}[B_1] = \text{tr}[I] - \text{tr}\left[\tfrac{1}{n}J_n\right] = n - 1$$

We can now use the result of Cochran's thm:

$$\sum(Z_i - \bar{Z})^2 \sim \chi^2_{n-1} \quad \text{indep of} \quad n\bar{Z}^2 \sim \chi^2_1$$
$$\bar{Z} = \frac{Z_1 + \ldots + Z_n}{n} = \frac{\frac{X_1-\mu}{\sigma} + \ldots + \frac{X_n-\mu}{\sigma}}{n} = \frac{X_1 + \ldots + X_n - n\mu}{n\sigma} = \frac{\cancel{n}\bar{X} - \cancel{n}\mu}{\cancel{n}\sigma} = \frac{\bar{X}-\mu}{\sigma}$$
$$\sum(Z_i - \bar{Z})^2 = \sum\left(\frac{X_i - \cancel{\mu}}{\sigma} - \frac{\bar{X} - \cancel{\mu}}{\sigma}\right)^2 = \frac{\sum(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)}{\sigma^2}S^2 \sim \chi^2_{n-1}$$
$$\Downarrow$$
$$S^2 \sim \text{Gamma}\left(\tfrac{n-1}{2}, \tfrac{n-1}{2\sigma^2}\right)$$
$$n\bar{Z}^2 = n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 = \frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2_1$$
$$\Rightarrow \bar{X}, S^2 \text{ are indep.}$$

This fact is only true for the iid normal rv. In every other rv the average and the sample variance are *dependent* (proved in 1936 by Geary).

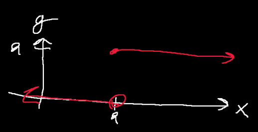
$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1), \quad \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim ?$$

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} = \frac{\bar{X}-\mu}{\frac{1}{\sqrt{n}}\sqrt{S^2}} = \frac{\bar{X}-\mu}{\frac{1}{\sqrt{n}}\sqrt{\frac{\sigma^2}{n-1}\frac{n-1}{\sigma^2}S^2}}$$
$$= \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{\frac{n-1}{\sigma^2}S^2}{n-1}}} = \frac{\overbrace{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}^{Z \sim N(0,1)}}{\underbrace{\sqrt{\frac{\frac{n-1}{\sigma^2}S^2}{n-1}}}_{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}} \sim T_{n-1}$$

Hodgepodge of theoretical results in probability

Let X be a non-neg rv which means Supp[X] ≥ 0 and has finite expectation. Consider:

$g(X) = a\mathbb{1}_{X \geq a}$ for $a > 0$

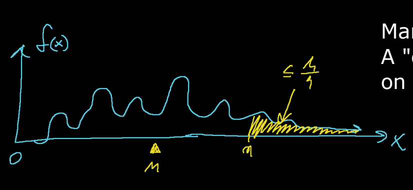


$$a\mathbb{1}_{X \geq a} \stackrel{?}{\leq} X$$

If $X < a$: l.h.s. $= 0$, r.h.s. $= X$, $0 \leq X$ ✓

If $X \geq a$: l.h.s. $= a$, r.h.s. $= X$, $a \leq X$ ✓

$$\Rightarrow E\left[a\mathbb{1}_{X \geq a}\right] \leq E[X]$$
$$\Rightarrow a\,E\left[\mathbb{1}_{X \geq a}\right] \leq \mu \qquad \mathbb{1}_{X \geq a} \sim \text{Bern}(P(X \geq a))$$
$$\Rightarrow a\,P(X \geq a) \leq \mu \Rightarrow P(X \geq a) \leq \frac{\mu}{a}$$



Markov's Inequality
A "crude" (inexact) bound on the right tail probability

$$X \sim \text{Exp}(1) := e^{-x}\mathbb{1}_{x \geq 0} \qquad P(X \geq x) = e^{-x}, \quad E[X] = 1$$
$$P(X \geq 1) = e^{-1} \approx 0.37 \qquad P(X \geq 2) = e^{-2} = .1353$$

Markov: "crude"

$$P(X \geq 1) \leq \tfrac{1}{1} = 1 \qquad P(X \leq 2) \leq \tfrac{1}{2} = 0.5$$

There are tons of corollaries of Markov's Inequality

- let $b = a\mu$ $\quad P(X \geq b) \leq \frac{\mu}{b} \Rightarrow P(X \geq a\mu) \leq \frac{1}{a}$
- let h(X) be a monotonically increasing function (thus 1:1) and Supp[h(X)] > 0

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

- let a = Quantile[X, p] $\stackrel{\text{if } X \text{ cont}}{=} F_X^{-1}(p)$

$$P(X \geq a) \leq \frac{\mu}{a}$$
$$\Rightarrow 1 - F_X(a) \leq \frac{\mu}{a}$$
$$\Rightarrow 1 - F_X(F_X^{-1}(p)) \leq \frac{\mu}{F_X^{-1}(p)}$$
$$\Rightarrow 1 - p \leq \frac{\mu}{F_X^{-1}(p)}$$
$$\Rightarrow F_X^{-1}(p) \leq \frac{\mu}{1-p}$$

we have an upper bound for a quantile

e.g Quantile[X, 90%] ≤ 10μ

- Let X be any rv with finite μ. Then |X| has non-neg support and finite μ.

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

- Let X be any rv with finite μ and finite σ^2. Let Y := $(X - \mu)^2$ Y is clearly a non-neg rv and it has finite expectation since $E[Y] = E[(X - \mu)^2] = \sigma^2$ which was assumed finite.

$$\Rightarrow P(Y \geq a^2) \leq \frac{E[Y]}{a^2}$$
$$\Rightarrow P((X-\mu)^2 \geq a^2) \leq \frac{E[(X-\mu)^2]}{a^2}$$
$$\Rightarrow P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Chebyshev's Inequality

Let's make a corollary for non-neg rv X and a ≥ μ

$$P(|X-\mu| \geq a) = P(X-\mu \geq a) + P(X - \mu \leq -a)$$
$$= P(X \geq \mu + a) + \underbrace{P(X \leq \mu - a)}_{Q \geq 0} \leq \frac{\sigma^2}{a^2}$$
$$= P(X \geq \mu + a) \leq \frac{\sigma^2}{a^2} - Q$$
$$\Rightarrow P(X \geq \mu + a) \leq \frac{\sigma^2}{a^2}$$

let $b = \mu + a$

$$\Rightarrow \boxed{P(X \geq b) \leq \frac{\sigma^2}{(b-\mu)^2}}$$

- Let X be any rv with finite μ and consider Y = e^{tX} which is a non-neg rv which hopefully has a finite expectation.

$$P(Y \geq c) \leq \frac{E[Y]}{c}$$