

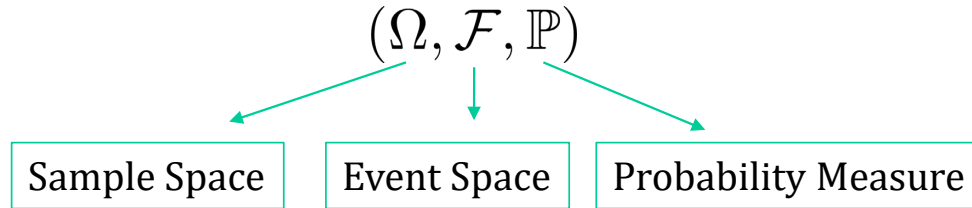
LEC017 Review of Probability I

VG441 SS2020

Cong Shi
Industrial & Operations Engineering
University of Michigan

Probability Space

- Probability space as a triplet



- A coin toss

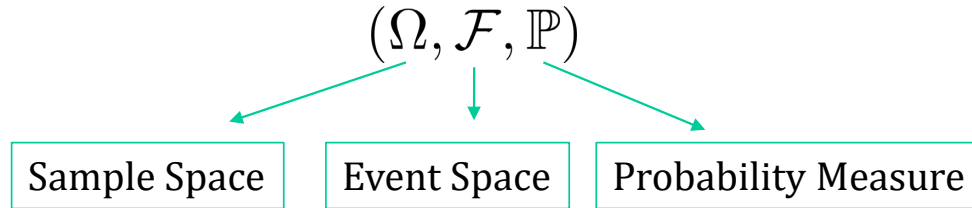
$$\Omega = \{H, T\}$$

$$\mathcal{F} = 2^\Omega = \{\emptyset, \{H\}, \{T\}, \Omega\}$$

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$$

Probability Axioms

- Probability space as a triplet



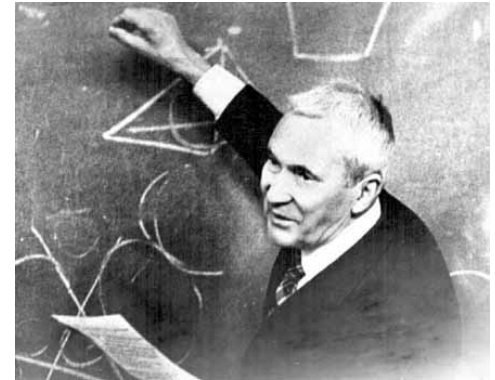
- 3 Axioms

$$\mathbb{P}(\Omega) = 1$$

If $A \in \mathcal{F}$, then $0 \leq \mathbb{P}(A) \leq 1$

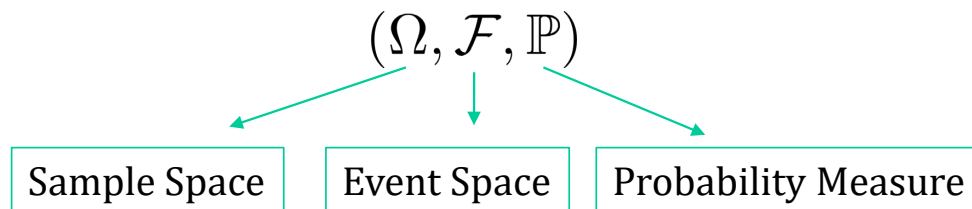
If $A_1, A_2, \dots \in \mathcal{F}$ and disjoint, then
 $\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$

Andrey Kolmogorov



Random Variable (RV)

- A convenient representation of sample space



- Random variable

$$X : \Omega \rightarrow \mathbb{R}$$

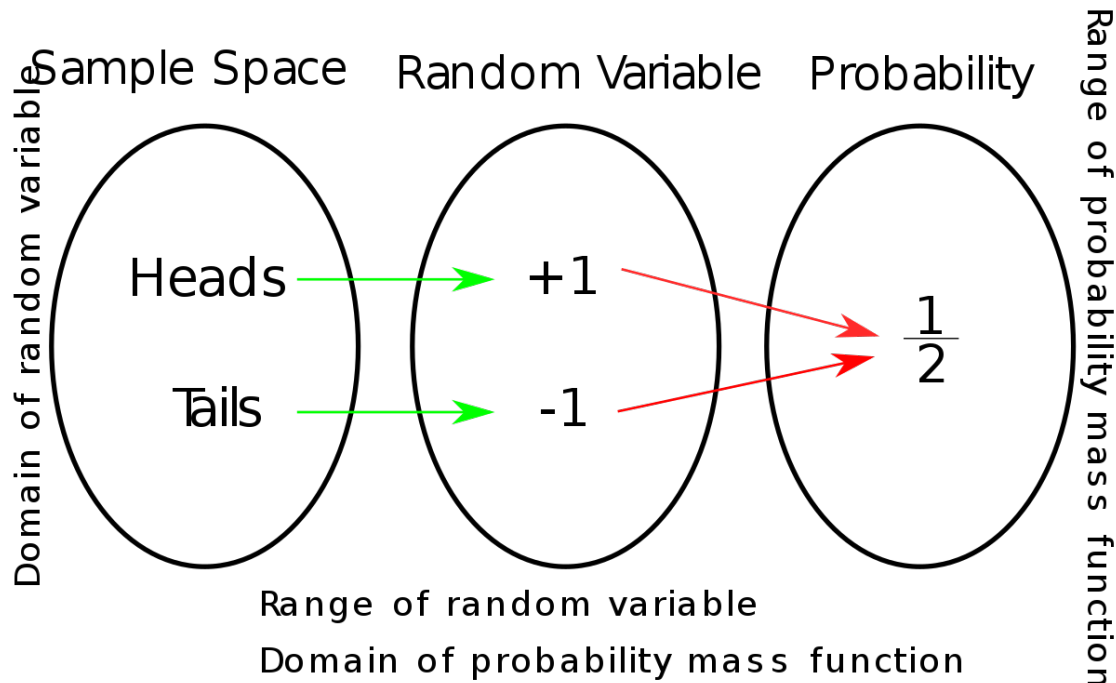
- Coin toss

$$X = \begin{cases} 0 & \text{if T} \\ 1 & \text{if H} \end{cases}$$

- Assign probability measure or distribution...

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p$$

A Fair Coin Toss (as Bernoulli($p=1/2$))



Common Random Variable (RV)

Discrete (PMF, CDF, MGF)

- Bernoulli
- Binomial
- Poisson
- Geometric
- Negative Binomial
- Hypergeometric
- Dirichlet (multinomial)
-

Continuous (PDF, CDF, MGF)

- Uniform
- Normal
- Exponential
- Beta
- Gamma
- Weibull
- Gumbel (extreme value)
-

PDF, CDF, Expectation, Moment, MGF

- Probability mass function (discrete) $\mathbb{P}(X = x)$
- Probability density function (continuous)

$\delta \rightarrow \bigcirc$ $f(x) \approx \frac{1}{\delta} \mathbb{P}(x \leq X \leq x + \delta)$

- Cumulative distribution function $F(x) = \mathbb{P}(X \leq x)$
- Expectation $\mathbb{E}(X)$
- Variance $\text{Var}(X)$
- Covariance $\text{Cov}(X, Y)$
- Correlation Coefficient $\rho_{X,Y}$
- Moments $m_k = \mathbb{E}(X^k)$
- Moment generating function $M(\theta) = \mathbb{E}(e^{\theta X})$

any para. GFR.

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx \\ &= \int_{-\infty}^{\infty} \left(1 + tx + \frac{t^2 x^2}{2!} + \cdots + \frac{t^n x^n}{n!} + \cdots \right) f(x) dx \\ &= 1 + tm_1 + \frac{t^2 m_2}{2!} + \cdots + \frac{t^n m_n}{n!} + \cdots, \end{aligned}$$

Conditional Probability

- Conditional probability

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- Bayes Theorem (A_i disjoint cases)

$$P(A_i | B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(B | A_i) P(A_i)}{P(B)} = \frac{P(B | A_i) P(A_i)}{\sum_{j=1}^n P(B | A_j) P(A_j)}$$

label data

prior
|
 $P(\text{label})$

like likelihood

$P(\text{data} | \text{label})$

posterior
|
 $P(\text{label} | \text{data})$

Conditional Probability

- Conditional probability

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- Bayes Theorem (A_i disjoint cases)

$$P(A_i \mid B) = \frac{P(B \mid A_i) P(A_i)}{P(B)} = \frac{P(B \mid A_i) P(A_i)}{\sum_{j=1}^n P(B \mid A_j) P(A_j)}$$

- What is the probability of a customer coming from Copenhagen spends above the median of the rest of the customers, on some item? Facts:
 - People from Copenhagen spent 19.5%, people from Hongkong spent 7.8% and the rest (of the world) spent 72.7%.
 - 48.4% of people from Copenhagen spent above the median
 - 35.2% of people from Hongkong spent above the median
 - 56.7% of people from the rest of the world spent above the median

17.67% 先算 prior 和 likelihood,
↓ Bayesian updating
得出 $P(A_i \mid B)$

Two Basic Inequalities

- Markov inequality

Let X be a *non-negative* r.v. Fix a constant $a > 0$. Then

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}$$

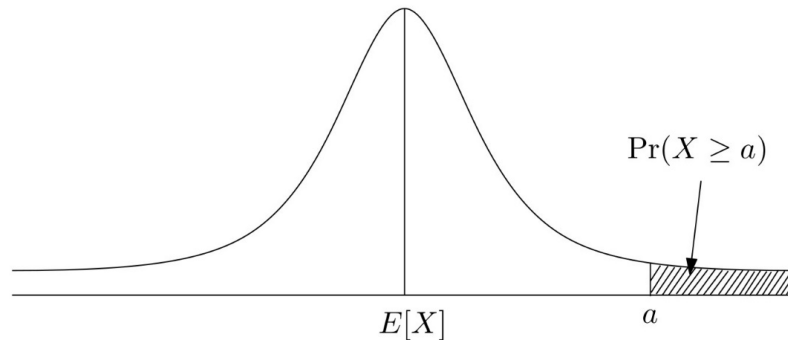


Figure : Markov's Inequality bounds the probability of the shaded region.

Two Basic Inequalities

- Markov inequality

Let X be a *non-negative* r.v. Fix a constant $a > 0$. Then

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}$$

1st moment

Proof. Define Y by

$$Y = \begin{cases} a & \text{if } X \geq a \\ 0 & \text{if } X < a \end{cases}$$

truncation of X

As $X \geq Y$ a.s., it follows that $\mathbb{E}(X) \geq \mathbb{E}(Y) = a\mathbb{P}(X > a)$. □

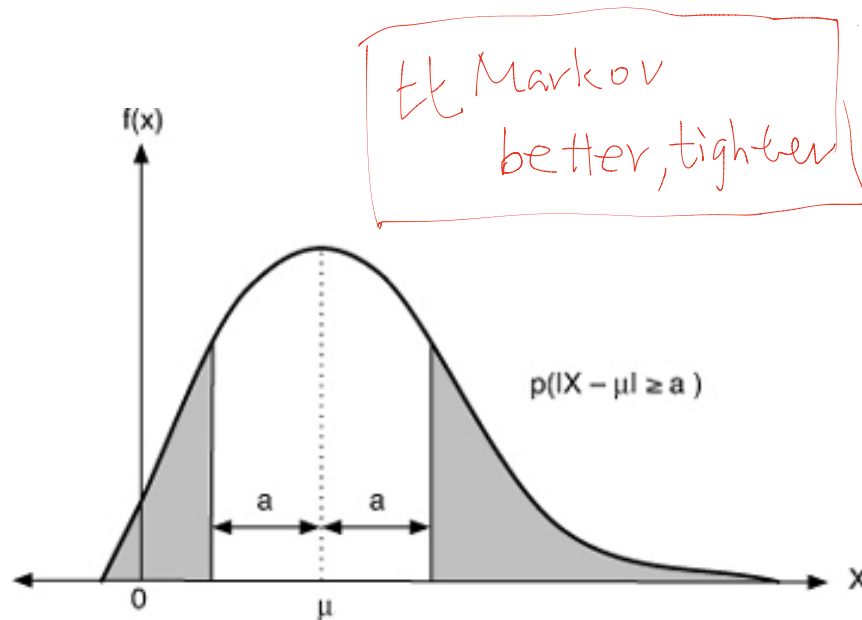
*almost surely
(along every possibility relation)*

Two Basic Inequalities

- Chebyshev's inequality

Let X be a r.v. having finite mean μ and variance σ^2 and $\epsilon > 0$ then

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



Two Basic Inequalities

- Chebyshev's inequality

Let X be a r.v. having finite mean μ and variance σ^2 and $\epsilon > 0$ then

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

Proof. Note that $(X - \mu)^2$ is a non-negative r.v. and

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

平方一下,

再用 Markov

□

Convergence of random variables

- Convergence in probability

Convergence in probability: $X_n \rightarrow X$ i.p. if for all $\epsilon > 0$, we have

$$\mathbb{P}(|X_n - X| \geq \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

- Convergence in distribution

Convergence in distribution (or weak convergence): Let X and X_n , $n \in \mathbb{N}$, be random variables with CDFs F and F_n , respectively. We say $X_n \xrightarrow{d} X$ if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x), \text{ for every } x \in \mathbb{R} \text{ at which } F \text{ is continuous.}$$

Suffice to show convergence in MGF, i.e., $M_n(\theta) \rightarrow M(\theta)$

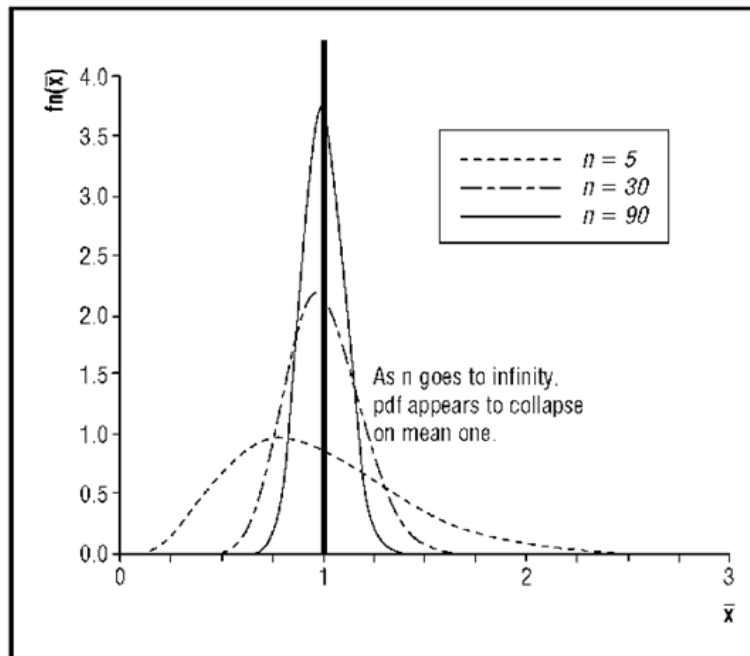


Limit Theorems

- (Weak) Law of Large Numbers

Let X_1, \dots, X_n be i.i.d. having finite mean $\mathbb{E}[X] = \mu$ and variance σ^2 , then

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{i.p.} \mathbb{E}[X] \quad \text{as } n \rightarrow \infty.$$



Limit Theorems

- (Weak) Law of Large Numbers

Let X_1, \dots, X_n be i.i.d. having finite mean $\mathbb{E}[X] = \mu$ and variance σ^2 , then

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \xrightarrow{i.p.} \quad \mathbb{E}[X] \quad \text{as } n \rightarrow \infty.$$

Proof. Since $\mathbb{E}(\bar{X}_n) = \mu$ and $\mathbf{Var}(\bar{X}_n) = \sigma^2/n$, by Chebyshev's inequality,

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathbf{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 n}.$$

Then we drive $n \rightarrow \infty$. □

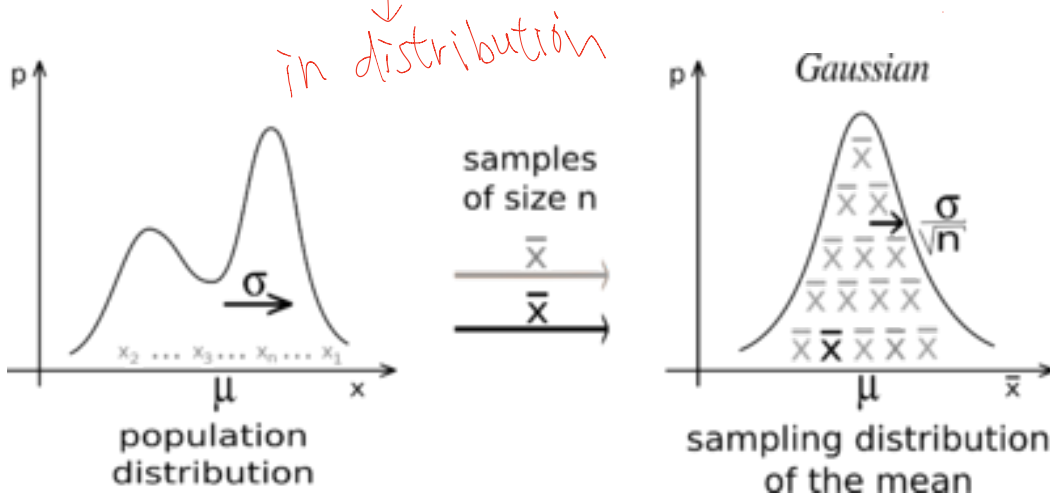
邦切比雪夫不等式

Limit Theorems

- Central Limit Theorem

Let $\{X_n\}$ be a sequence of i.i.d. r.v.'s with $\mathbb{E}[X_1] = \mu$ and $\mathbf{Var}(X_1) = \sigma^2 < \infty$.
 Let $S_n = \sum_{k=1}^n X_k$. Then

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d} \sigma N(0, 1) \quad \text{as } n \rightarrow \infty.$$



$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

i.e. $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$

Limit Theorems

- MGF of $N(\mu, \sigma^2)$

自学一波

Normal density

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

For standard normal Z:

$$\begin{aligned} M_Z(\theta) &= \mathbb{E}[e^{\theta Z}] = \int_{-\infty}^{\infty} e^{\theta x} f_Z(x) dx \\ &= \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2x\theta)} dx \\ &= e^{\theta^2/2} \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}}_{\text{pdf of } \mathcal{N}(\theta, 1)} dx \\ &= e^{\theta^2/2} \end{aligned}$$

For any normal X:

$$X = \mu + \sigma Z$$

$$\begin{aligned} M_X(\theta) &= \mathbb{E}[e^{\theta(\mu + \sigma Z)}] \\ &= e^{\theta\mu} \mathbb{E}[e^{\theta\sigma Z}] \\ &= e^{\theta\mu} M_Z(\theta\sigma) \\ &= e^{\theta\mu} e^{\theta^2\sigma^2/2} \\ &= e^{(\mu\theta + \frac{\sigma^2\theta^2}{2})} \end{aligned}$$

$\mu=0$ 时 $M_X(\theta) = e^{\frac{\sigma^2\theta^2}{2}}$

Limit Theorems

• Central Limit Theorem

Let $\{X_n\}$ be a sequence of i.i.d. r.v.'s with $\mathbb{E}[X_1] = \mu$ and $\mathbf{Var}(X_1) = \sigma^2 < \infty$.
Let $S_n = \sum_{k=1}^n X_k$. Then

generally without loss of generality

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d} \sigma N(0, 1) \quad \text{as } n \rightarrow \infty.$$

Proof. WLOG, assume $\mathbb{E}[X_1] = \mu = 0$. It suffices to show that the MGF of S_n/\sqrt{n} converges to that of $Z = N(0, \sigma^2)$. Let $M(\theta) := \mathbb{E}[e^{\theta X_1}]$. Then we have $M'(0) = \mathbb{E}[X_1] = 0$ and $M''(0) = \mathbb{E}[X_1^2] = \sigma^2$. For each $\theta \in \mathbb{R}$,

中间步骤

$$\begin{aligned} \mathbb{E}\left[e^{\theta \frac{S_n}{\sqrt{n}}}\right] &= \left(M\left(\frac{\theta}{\sqrt{n}}\right)\right)^n \stackrel{\text{泰勒}}{=} \left(1 + \frac{M''(0)}{2} \frac{\theta^2}{n} + o\left(\frac{\theta^2}{n}\right)\right)^n \\ &= \left(1 + \frac{\sigma^2 \theta^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n \xrightarrow{n \rightarrow \infty} e^{\frac{\sigma^2 \theta^2}{2}}, \end{aligned}$$

$\mathbb{E}[e^{\theta(X_1 + \dots + X_n)/\sqrt{n}}]$
 $= \mathbb{E}[e^{\theta/\sqrt{n} X_1}] \cdot \mathbb{E}[e^{\theta/\sqrt{n} X_2}] \dots$

which is the MGF of $N(0, \sigma^2)$.

试图 $M_{S_n/\sqrt{n}}(\theta) \rightarrow M_{N(0, \sigma^2)}(\theta) = e^{\frac{\sigma^2 \theta^2}{2}}$

□