LEC017 Review of Probability I

VG441 SS2021

Cong Shi
Industrial & Operations Engineering
University of Michigan

Probability Space

- Probability space as a triplet

$$(\Omega, \mathcal{F}, \mathbb{P})$$

Sample Space | Event Space | Probability Measure

- A coin toss

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

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

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

Probability Axioms

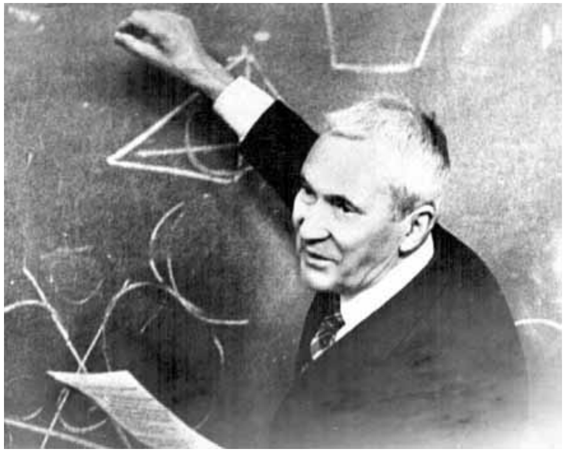
- Probability space as a triplet

$$(\Omega, \mathcal{F}, \mathbb{P})$$

Sample Space | Event Space | Probability Measure

- 3 Axioms

Andrey Kolmogorov



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

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

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

Random Variable (RV)

- A convenient representation of sample space

$$(\Omega, \mathcal{F}, \mathbb{P})$$

$$\boxed{\text{Sample Space}} \quad \boxed{\text{Event Space}} \quad \boxed{\text{Probability Measure}}$$

- Random variable

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

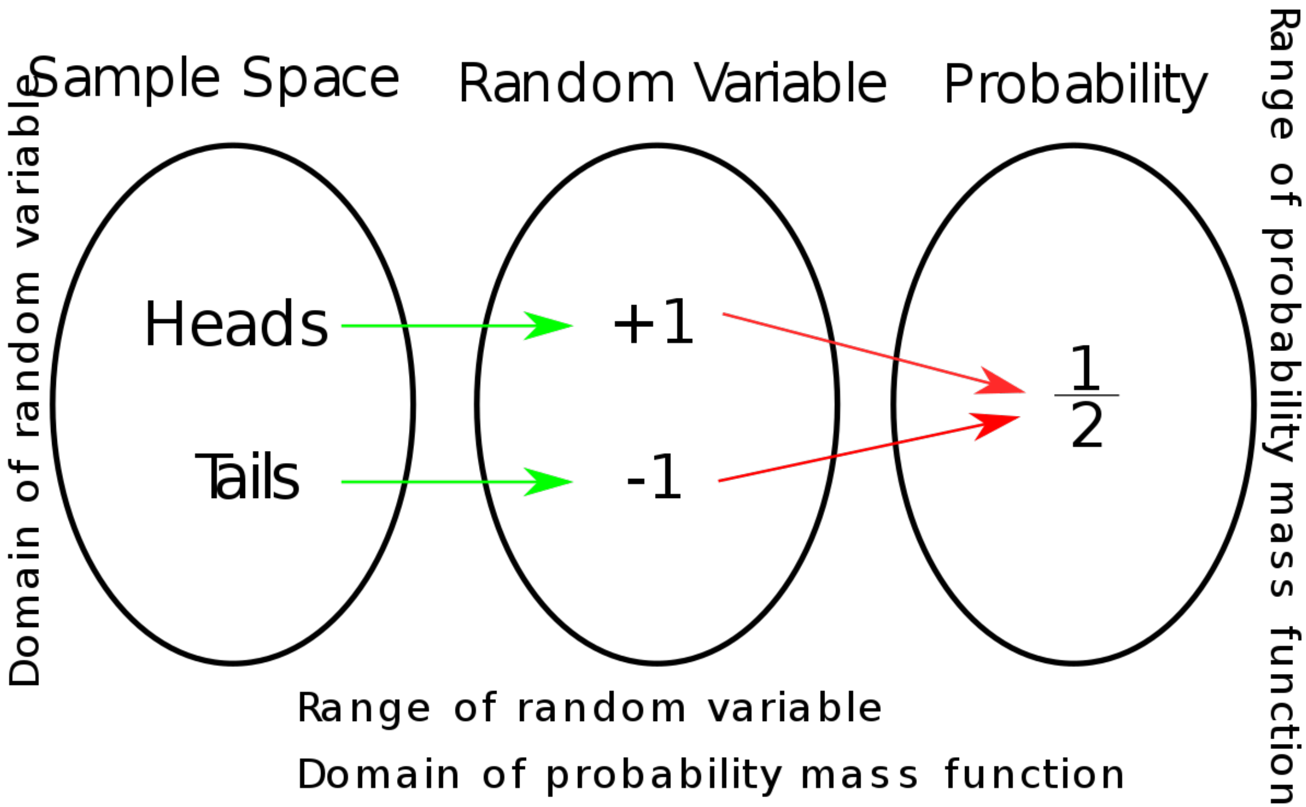
- Coin toss

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

- 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)

$$f(x) \approx \tfrac{1}{\delta}\mathbb{P}(x \le X \le x + \delta)$$

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

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

Conditional Probability

- Conditional probability

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

- Bayes Theorem (A_i disjoint cases)

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

Conditional Probability

- Conditional probability

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

- Bayes Theorem (A_i disjoint cases)

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

- What is the probability of a customer coming from Copenhagen given that she spends above the median? Facts:
  - People from Copenhagen 19.5%, people from Hongkong 7.8% and the rest (of the world) 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

Two Basic Inequalities

- Markov inequality

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

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

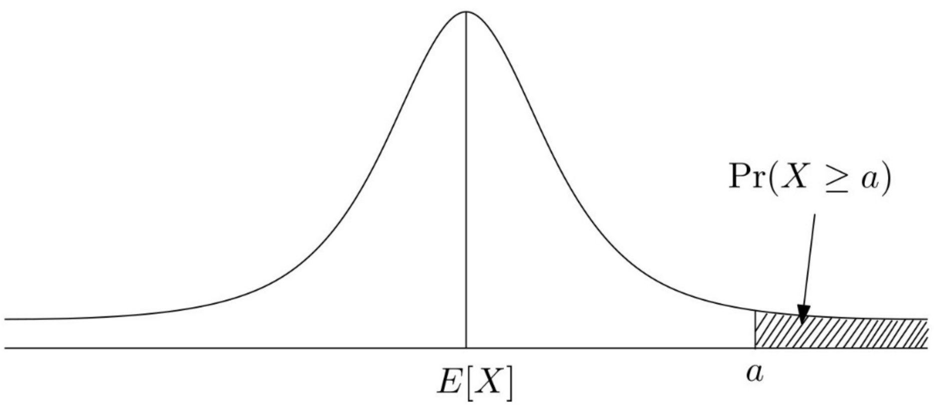


Figure : Markov's Inequality bounds the probability of the shaded region.

Two Basic Inequalities

- Markov inequality

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

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

Proof. Define Y by

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

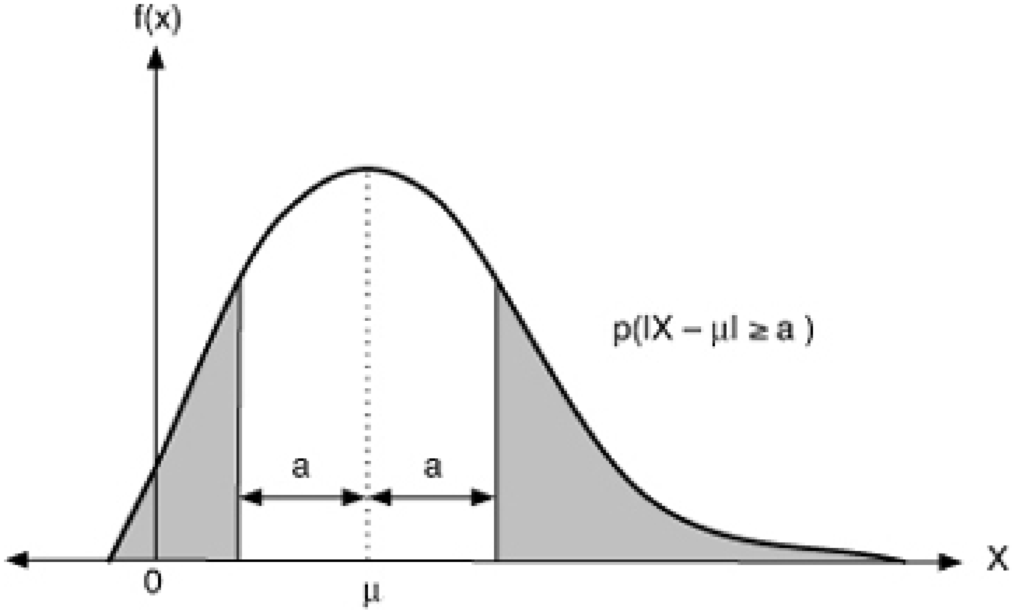
As $X \geq Y$ a.s., it follows that $\mathbb{E}(X) \geq \mathbb{E}(Y) = a\mathbb{P}(X > a)$. $\square$

Two Basic Inequalities

- Chebyshev's inequality

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

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



Two Basic Inequalities

- Chebyshev's inequality

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

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

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

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

$\square$

Convergence of random variables

- Convergence in probability

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

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

- 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

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

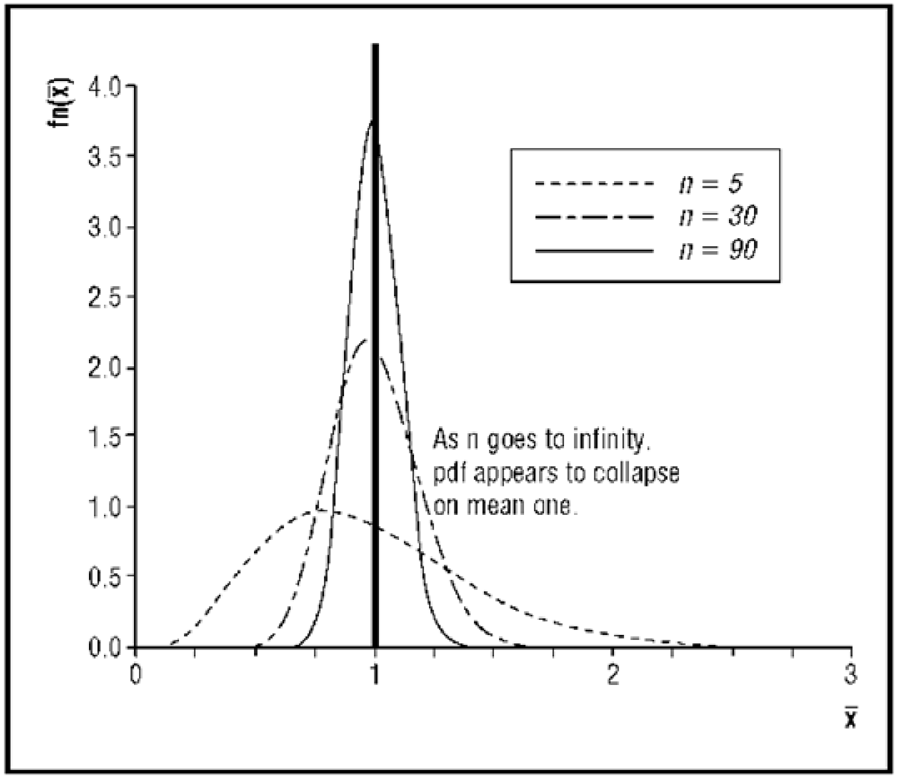
Suffice to show convergence in MGF, i.e., $M_n(\theta)$ ⍰ $M(\theta)$

Limit Theorems

- (Weak) Law of Large Numbers

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

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



- (Weak) Law of Large Numbers

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

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

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

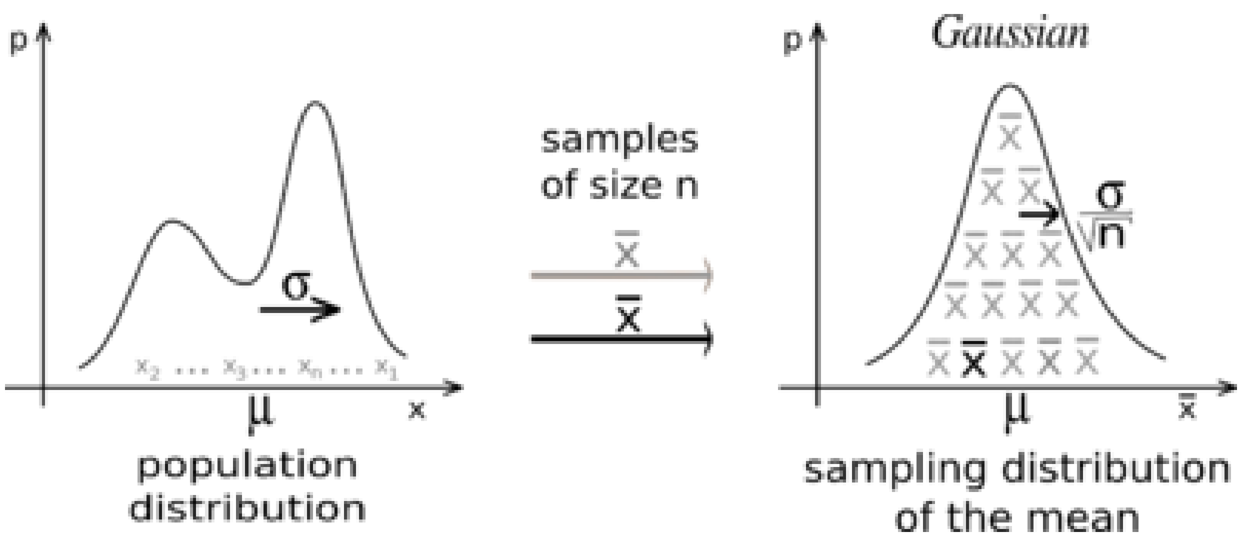
$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| \geq \epsilon\right) \leq \frac{\mathbf{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 n}.$$

Then we drive $n \to \infty$. $\square$

- Central Limit Theorem

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

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



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

Limit Theorems

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

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

For standard normal Z:

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

For any normal X:

$$\boxed{X = \mu + \sigma Z}$$

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

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

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

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 $M(\theta) := \mathbb{E}\left[e^{\theta X_1}\right]$. Then we have $M'(0) = \mathbb{E}\left[X_1\right] = 0$ and $M''(0) = \mathbb{E}\left[X_1^2\right] = \sigma^2$. For each $\theta \in \mathbb{R}$,

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

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

$\square$