MAS250 Probability and Statistics

CHAPTER 4

RANDOM VARIABLES AND EXPECTATION

4.1 Random Variables

$X := \#$ of heads

$$S = \begin{pmatrix} HH \\ HT \\ TH \\ TT \end{pmatrix} \Rightarrow X = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{cases} P\{X=0\} = \frac{1}{4} \\ P\{X=1\} = \frac{1}{2} \\ P\{X=2\} = \frac{1}{4} \end{cases}$$

- Consider an experiment

- Random variables : the quantities of interest that are determined by the result of the experiment

- Example
  - X is the sum of two fair dice

Distribution Function

- The (cumulative) distribution function CDF of a random variable X

$$F(x) \coloneqq P\{X \leq x\}$$

$$P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\}$$
$$= F(b) - F(a)$$

- Notation
  - $X \sim F(x)$
  - We sometimes use $F_X(x)$ instead of $F(x)$ for r.v. X.

4.2 Types of random variables

- Discrete random variable X
  - The set of possible values of X is countable
  - The probability mass function (pmf) $p(x)$

$$p(x) := P\{X = x\}$$

- Let X take its values on $\{x_1, x_2, \cdots\}$
  - $F(x) \overset{:=P\{X\leq x\}}{=} \sum_{x_i \leq x} p(x_i)$
  - $p(x_i) > 0, i = 1,2, \cdots$, and
  - $p(x) = 0$ for all other values of x
  - $\sum_{i=1}^{\infty} p(x_i) = 1$

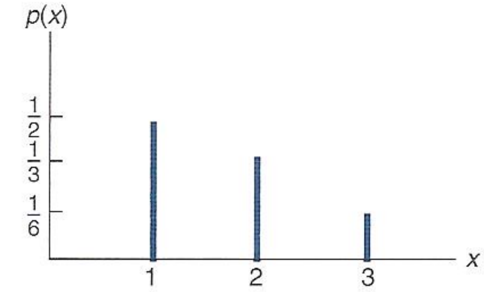


FIGURE 4.1 *Graph of (p)x, Example 4.2a.*

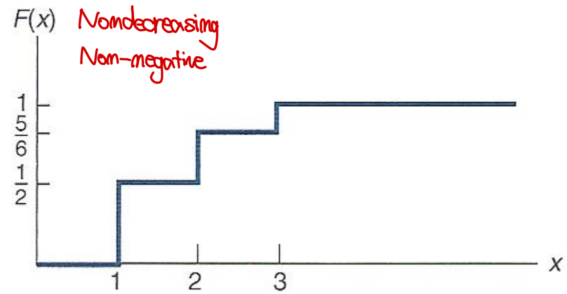


FIGURE 4.2 *Graph of F(x).*

- Continuous random variables
  - X is a continuous random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set of real numbers

$$P\{X \in B\} = \int_B f(x)dx$$

  - $f(x)$ is called the probability density function (pdf) of the random variable X. Note that

$$\int_{-\infty}^{\infty} f(x)dx = P\{X \in R\} = 1$$

- The distribution function $F(x)$ of X

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

$F(x) = 0.5$
$\Rightarrow x$ is median

- Properties of the pdf
  - $P\{a < X \leq b\} = \int_a^b f(x)dx$
  - $P\{X = a\} = \int_a^a f(x)\,dx = 0$
  - $\frac{d}{dx}F(x) = f(x)$

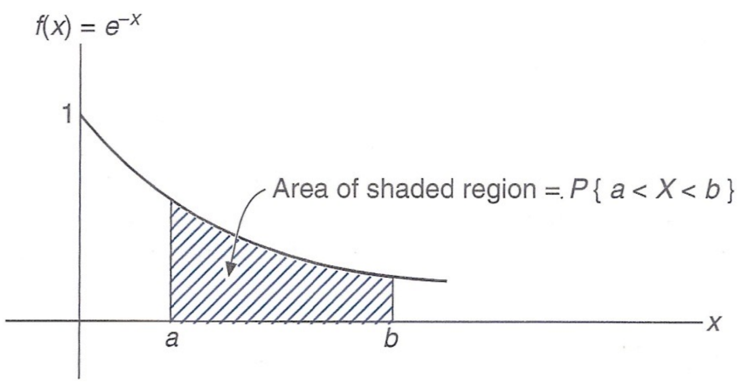


FIGURE 4.3 *The probability density function* $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$.

$$P\left\{a - \frac{\epsilon}{2} < X \leq a + \frac{\varepsilon}{2}\right\} = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\epsilon}{2}} f(x)dx \approx f(a)\varepsilon$$

- Example 4.2a

Consider a discrete r.v. X that takes its values on $\{1,2,3\}$. If we know that $p(1) = \frac{1}{2}, p(2) = \frac{1}{3}$, what is $p(3)$? $= 1 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$.

- Example 4.2b

  Suppose that X is a continuous r.v. whose pdf is given by

$$f(x) = \begin{cases} C(4x - 2x^2), & 0 < x < 2 \\ 0, & otherwise \end{cases}$$

(a) What is the value of C?

(b) Find $P\{X > 1\}$.

$C\int_0^2 4x - 2x^2\,dx = -\frac{2}{3}x^3 + 2x^2 \Big|_0^2 \cdot C = 1$

$\therefore C = \frac{3}{8}$ ∎

$C\int_1^2 4x - 2x^2\,dx = \frac{1}{2}$ ∎

4.3 Jointly distributed RVs

- The joint (cumulative) distribution function of random variables X and Y
$$F(x,y) \coloneqq P\{X \le x, Y \le y\}$$

- The marginal distribution functions of X and Y

$$F(x,\infty) = P\{X \le x, Y \le \infty\} = P\{X \le x\} = F_X(x)$$
$$F(\infty,y) = P\{X \le \infty, Y \le y\} = P\{Y \le y\} = F_Y(y)$$

- When X and Y are discrete r.v.s, the joint pmf of X and Y is given by

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

- X and Y are said to be jointly continuous if there exists a function $f(x, y)$ (called a joint pdf) defined for all real numbers such that for every set $C \subset R^2$

$$P\{(X, Y) \in C\} = \iint_C f(x, y)\, dxdy$$

- The relation between joint pmf and marginal pmf
  - $p_X(x_i) = P\{X = x_i\} = \sum_{j=1}^{\infty} p(x_i, y_j)$
  - $p_Y(y_j) = P\{Y = y_j\} = \sum_{i=1}^{\infty} p(x_i, y_j)$
- The relation between joint pdf and marginal pdf
  - $P\{X \le x\} = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(t, s)\, ds\, dt = \int_{-\infty}^{x} f_X(t) dt$
  - $P\{Y \le y\} = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(t, s)\, dt\, ds = \int_{-\infty}^{y} f_Y(s) ds$

- Properties of the joint pdf $f(x, y)$
  - $P\{X \in A, Y \in B\} = \int_B \int_A f(x, y) dx dy$
  - $F(a, b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y) dx dy$
  - $f(a, b) = \frac{\partial^2}{\partial x \partial y} F(x, y)|_{(x,y)=(a,b)}$
  - $P\{a < X \leq a + da, b < Y \leq b + db\}$

    $= \int_a^{a+da} \int_b^{b+db} f(x, y) dy\, dx \cong f(a, b) da db$

- Example 4.3c

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x, y < \infty \\ 0, & otherwise \end{cases}$$

  - Compute $P\{X > 1, Y < 1\}$.
  - Compute $P\{X < Y\}$.
  - Compute $P\{X < a\}$.

① $= \int_0^1 \int_1^\infty 2e^{-x} e^{-2y}\, dx\,dy$

$= \int_0^1 e^{-2y} \left[-2e^{-x}\right]_1^\infty dy$

$= \int_0^1 2e^{-1} e^{-2y}\, dy$

$= -e^{-1}e^{-2y} \Big|_0^1 = e^{-1}(-e^{-2}+1)$ ■

② $\int_0^\infty \int_0^y 2e^{-x}e^{-2y}\, dx\,dy \quad \left(= \int_0^\infty \int_x^\infty \right.$

$= \int_0^\infty 2e^{-2y}\left[-e^{-x}\right]_0^y dy$

$= \int_0^\infty 2e^{-2y} dy - \int_0^\infty 2e^{-3y} dy$

$= 1 - \frac{2}{3} = \frac{1}{3}$ ■

③ $\int_0^\infty \int_0^a f(x,y)\, dx\,dy$

$= 1 - e^{-a}$ ■

Independent r.v.s

$f(x,y) = 2e^{-x}e^{-2y} = 2e^{-x} \times e^{-2y} = F_X(x) \cdot F_Y(y) \Rightarrow$ Indep.

- X and Y are said to be independent if for any two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\} P\{Y \in B\}$$

$$\Leftrightarrow F(x,y) = F_X(x)F_Y(y) \quad \forall x,y \in R$$

  - When X and Y are discrete, for all x and y

$$p(x,y) = p_X(x)p_Y(y)$$

  - When X and Y are jointly continuous, for all x and y

$$f(x,y) = f_X(x)f_Y(y)$$

- Example 4.3d

  Suppose X and Y are independent r.v.s having the same common pdf

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & otherwise. \end{cases}$$

  Find the pdf of the r.v. $\frac{X}{Y}$.

CDF:

$$F_{\frac{X}{Y}}(a) = P\left(\frac{X}{Y} \leq a\right) = P(X \leq Ya)$$

$$= \int_0^\infty \int_0^{ay} e^{-x} e^{-y}\, dx\, dy$$

$$= \int_0^\infty e^{-y} \left[-e^{-x}\right]_0^{ay} dy$$

$$= \int_0^\infty e^{-y} - e^{-(a+1)y}\, dy = 1 - \frac{1}{a+1}.$$

PDF:

$$f_{\frac{X}{Y}}(a) = F'_{\frac{X}{Y}}(a) = \frac{1}{(a+1)^2} \quad (0 < a < \infty)$$

- For r.v.s $X_1, X_2, \cdots, X_n$, their joint probability distribution function is defined by

$$F(x_1, \cdots, x_n) \coloneqq P\{X_1 \le x_1, \cdots, X_n \le x_n\}$$

- The joint pmf

$$p(x_1, \cdots, x_n) \coloneqq P\{X_1 = x_1, \cdots, X_n = x_n\}$$

- The joint pdf (jointly continuous), $C \in R^n$

$$P\{(X_1, X_2, \cdots, X_n) \in C\} = \int \cdots \int_C f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

- For r.v.s $X_1, X_2, \cdots, X_n$, they are said to be independent if, for all $x_1, \cdots, x_n \in R$

$$P\{X_1 \leq x_1, \cdots, X_n \leq x_n\} = \prod_{i=1}^{n} P\{X_i \leq x_i\}$$

for identically distributed $:= \left(P\{X_i \leq x_i\}\right)^n$

Conditional distributions

- Let X and Y be discrete. The conditional pmf of X, given that $Y = y$, is defined by

$$p_{X|Y}(x|y) \coloneqq P\{X = x | Y = y\}$$

$$= \frac{P\{X=x, Y=y\}}{P\{Y=y\}}$$

$$= \frac{p(x,y)^{\text{joint pmf}}}{p_Y(y)_{\text{marginal pmf}}}, \quad p_Y(y) > 0$$

- Let X and Y be jointly continuous with $f(x, y)$. The conditional pdf of X, given that $Y = y$, is defined by

$$f_{X|Y}(x|y) := \frac{f(x,y)}{f_Y(y)}, \qquad f_Y(y) > 0$$

- $P\{x < X \le x + dx \mid y < Y \le y + dy\}$
$$\approx \frac{f(x,y)dxdy}{f_Y(y)dy} = f_{X|Y}(x|y)dx$$
- $P\{X \in A \mid Y = y\} = \int_A f_{X|Y}(x|y)\, dx$

$f(x,y) = f_{X|Y}(x)f_Y(y)$

- $P\left\{X \in A, Y \in B\right\} = \int_B P\{X \in A \,\Big|\, Y = y\}f_Y(y)dy$

$= \int_A \int_B f(x,y)dydx$

- c.f. $E[\,X \mid Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$

- Example 4.3h

  The joint pdf of X and Y is given by

$$f(x,y) = \begin{cases} \frac{12}{5}x(2-x-y), & 0 < x, y < 1, \\ 0, & otherwise. \end{cases}$$

compute $f_{X|Y}(x|y)$. $= \frac{f(x,y)}{f_Y(y)}$

$$f_Y(y) = \int_0^1 \frac{12}{5}x(2-x-y)\,dx$$

$$= \frac{12}{5}\left(\frac{2}{3} - \frac{y}{2}\right)$$

$$\therefore \frac{\frac{12}{5}x(2-x-y)}{\frac{12}{5}\left(\frac{2}{3} - \frac{y}{2}\right)} = \frac{6x(2-x-y)}{4-3y}$$

4.4 Expectation (mean)

Parameter

- X : a discrete r.v.

$$\mu = E[X] := \sum_i x_i P\{X = x_i\}$$

- X : a continuous r.v. with pdf $f(x)$

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

4.5 Properties of expectation

Let $g(x)$ be a real-valued function.

- X : a discrete r.v. with pmf $p(x)$

$$E[g(X)] := \sum_x g(x)p(x)$$

$E(X^2) = \sum x^2 p(x)$

- X : a continuous r.v. with pdf $f(x)$

$$E[g(X)] := \int_{-\infty}^{\infty} g(x)f(x)dx$$

- Example 4.5a

  a discrete r.v. X with pmf

  $$p(0) = 0.2, p(1) = 0.5, p(2) = 0.3.$$

  Compute $E[X^2]$. $= 0 \cdot 0.2 + 1 \cdot 0.5 + 4 \cdot 0.3$
  $= 1.7$.

- Example 4.5b

  a continuous r.v. X with pdf $f(x) = 1, 0 < x < 1.$

  Compute $E[X^3]$. $= \int_0^1 x^3 \cdot 1 \, dx = \frac{1}{4}$.

- For two constants a and b

$$E[aX + b] = aE[X] + b$$

- The nth moment of X

$n = 1 \Rightarrow$ mean
$n = 2 \Rightarrow$ variance
$n = 3 \Rightarrow$ skewness
$n = 4 \Rightarrow$ kurtosis

$$E[X^n] = \begin{cases} \sum_x x^n p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^n f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Expectation of the sum of r.v.s

- Let X and Y be discrete with joint pmf $p(x, y)$.

$$E[g(X,Y)] = \sum_y \sum_x g(x,y)\, p(x,y)$$

If $f(x.y) = f_X(x) f_Y(y)$,

$g(X.Y) = xy. \Rightarrow E[xy] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x.y) dx dy. \Rightarrow \int x f_X(x) dx \int y f_Y(y) dy = E(X) E(Y)$

- Let X and Y be jointly continuous with $f(x, y)$.

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy$$

$E[X+Y] = \iint (x+y) f(x,y)\,dx\,dy$

$= \iint x f(x,y)\,dx\,dy + \iint y f(x,y)\,dx\,dy$

$= \int x f_X(x)\,dx + \int y f_Y(y)\,dy = E(X) + E(Y)$

- When $g(x, y) = x + y$, we have

$$E[X + Y] = E[X] + E[Y]$$

(provided that both expectations exist.)

- In general, we have

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

$$X_i = \begin{cases} 1 & \text{if ith letter is placed correctly} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{i=1}^{N} X_i$$

$$E(X) = E(\sum X_i) = \sum E(X_i)$$

$$E(X_i) = 1 \cdot \frac{1}{N} + 0 \cdot (1 - \frac{1}{N}) = \frac{1}{N}$$

$$\therefore E(X) = 1$$

- Example 4.5g

  A secretary has typed N letters along with their respective envelopes. The envelopes get mixed up when they fall on the floor. If the letters are placed in the mixed-up envelopes in a completely random manner (that is, each letter is equally likely to end up in any of the envelopes), what is the expected number of letters that are placed in the correct envelopes?

$$X_i = \begin{cases} 1 & \text{at least one type } i \text{ coupon is contained in 10 coupons} \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{i=1}^{20} X_i$$

$$E(X) = \sum_{i=1}^{20} E(X_i)$$

$$p(X_i=1) = 1 - p(\text{no type } i \text{ in 10 coupon})$$

$$= 1 - \left(\frac{19}{20}\right)^{10}$$

$$E(X_i) = 1 - \left(\frac{19}{20}\right)^{10} \qquad \therefore E(X) = 20 - \frac{19^{10}}{20^9} \approx 8.025$$

- Example 4.5h

  Suppose there are 20 different types of coupons and suppose that each time one obtains a coupon it is equally likely to be any one of the types. Compute the expected number of different types that are contained in a set for 10 coupons.

The meaning of expectation

- When we predict X will equal c, the squared error is $(X-c)^2$.
  Considering $E[(X-c)^2]$, we see that the expectation $E[X]$ is the value that minimizes the mean square error $E[(X-c)^2]$.

$E[(x-c)^2] = E(x^2) - 2cE(x) + c^2.$

$= \{c - E(x)\}^2 + \underbrace{E(x^2) - \{E(x)\}^2}_{= var(x).}$

If $E[|x-c|]$, c is median.

4.6 Variance

Parameter

- Let X be a random variable with mean μ. The variance of X, denoted by $Var(X)$, is

$$Var(X) := E[(X-\mu)^2] = E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - \mu^2$$

We can also show that, for constants a and b

$$\sigma^2 = Var(X) = E[X^2] - \mu^2 \Rightarrow E[X^2] \geq \mu^2.$$

$$\sigma = Sd(X) = \sqrt{Var(X)}.$$

$$Var(aX+b) = a^2 Var(X)$$

4.7 Covariance

- Let X be a random variable with mean μ_X, and Y be a random variable with mean μ_Y.

The covariance of X and Y, $Cov(X, Y)$, is

$$Cov(X,Y) \coloneqq E[(X-\mu_X)(Y-\mu_Y)]$$

$= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X \mu_Y$

$= E(XY) - \mu_X \mu_Y$

We can also show that

$$Cov(X,Y) = E[XY] - \mu_X \mu_Y$$

Properties of covariance

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, X) = Var(X)$
- $Cov(aX, Y) = a\, Cov(X, Y)$
- $Cov(X + Z, Y) = Cov(X, Y) + Cov(Z, Y)$
- $Cov(\sum_{i=1}^{n} X_i, Y) = \sum_{i=1}^{n} Cov(X_i, Y)$
- $Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j)$
- $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} Cov(X_i, X_j)$
- If X and Y are independent, then $Cov(X, Y) = 0$.
- For independent r.v.s $\{X_i\}$, $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$

$Cov(X.Y) = E(XY) - \mu_X \mu_Y$

$= E(X)E(Y) - \mu_X \mu_Y = 0$ for indep. X.Y.

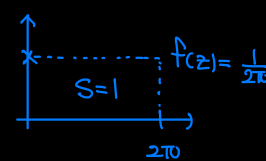
- Even though X and Y are uncorrelated, i.e., $Cov(X, Y) = 0$, X and Y may not be independent.

  Quadratic or etc.

  - Counterexample

    → Uniform distribution

    $X = \cos Z$, $Y = \sin Z$, $Z \sim U[0, 2\pi]$



    In this case, obviously X and Y are not independent because $X^2 + Y^2 = 1$.

    However, $E[X] = E[Y] = E[XY] = 0$, which implies that are uncorrelated.

    $\int_0^{2\pi} \cos z\, dz = \int_0^{2\pi} \sin z\, dz = \int_0^{2\pi} \sin z \cos z\, dz = 0.$

$\mu \cdot \sigma^2 \cdot \rho \Rightarrow$ parameter $\quad r = \hat{\rho} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})\sum(y-\bar{y})}}$

- Correlation coefficient of r.v.s X and Y

$$\rho = Corr(X,Y) := \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- Note that $|Corr(X,Y)| \leq 1$.

4.8 Moment Generating Function

- The moment generating function (mgf) $\phi(t)$ of a random variable X is defined by, $t \in R$,

$$\phi_X(t) = E[e^{tX}] = \begin{cases} \displaystyle\sum_x e^{tx} p(x) & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

- The mgf does not always exist for all t.

$f(x)$.

$E[X] = \int x f(x) dx$

$V[X] = \int (x-\mu)^2 f(x) dx$

$\Longleftrightarrow$

$\phi_X(t) = E[e^{tX}]$

$E(X) = \phi_X'(0)$

$E(X^2) = \phi_X''(0)$

$V(X) = E(X^2) - E(X)^2$

$= \phi_X''(0) - \phi_X'(0)^2$.

$E[Xe^{tX}]$

$\parallel$

- Note that $\phi_X{}'(t) = E\left[\frac{d}{dt} e^{tX}\right]$. So $\phi_X{}'(0) = E[X]$.
- Similarly, we have

$$\phi_X^{(n)}(0) = E[X^n], n \geq 1.$$

- For independent X and Y, we have

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

$= E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}] = E(e^{tX}) E(e^{tY})$

If X_i are ident. dist.,

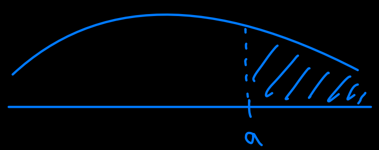
$\phi_{\sum_1^m X}(t) = \phi_X(t)^m$.

- The mgf uniquely determines the distribution.

4.9 Chebyshev's inequality

- Proposition 4.9.1 (Markov Inequality)

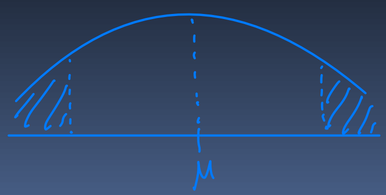
  For a nonnegative r.v. X and a constant $a > 0$,



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

- Proposition 4.9.2 (Chebyshev's Inequality)

  For a r.v. X with mean μ and variance σ^2, $k > 0$,



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

The Weak Law of Large Numbers

- Theorem 4.9.3 (The WLLN)

Let $X_1, X_2, \cdots$, be a sequence of independent and identically distributed (i.i.d.) r.v.s with mean μ.

$Var(X) = \sigma^2 < \infty$.

Then, for any $\epsilon > 0$,

$$P\left\{\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| > \epsilon\right\} \to 0 \text{ as } n \to \infty.$$