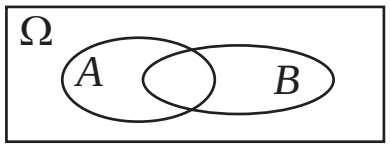
Lecture Notes 2
Review of Probability and Statistics

- Review of Probability
- Primer on Stochastic Processes
- Review of Hypothesis Testing
- Be sure to check course webpage for hw, announcements, etc.

Foundations of probability

- **Experiment** : outcome is unknown beforehand.
- **Sample space** Ω : set of all possible outcomes ω of experiment.
- **Event** : a subset of the sample space Ω
  - Event A **occurs** if the outcome $\omega \in A$.



For any two events A and B, we define

- **Union**:

$$A \cup B \;=\; \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$$

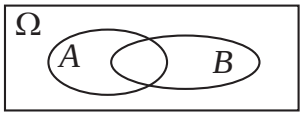
- **Intersection**:

$$AB \;\equiv\; A \cap B \;=\; \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$$

  - A, B **mutually exclusive** (or **disjoint**) if $A \cap B = \emptyset$.

- **Complement**:

$$\overline{A} \;=\; \{\omega \in \Omega : \omega \notin A\}$$



Example: Flip coin three times

- $\Omega \;=\; \{HHH,\; HHT,\; \ldots,\; TTT\}$.
- If outcome $\omega = HHT$, then heads on first and second tosses and tails on third toss.
- If event $A = \{H$ on second toss and T on third toss$\}$, then

$$A \;=\; \{HHT,\; THT\}.$$

Example: return on Intel stock (INTC) next week.

- Sample space: $\Omega = \mathbf{R}$
- If want to know if INTC increases at least 2%, then can let event $A = \{\%$ increase $\geq 2\}$

Example: Race with 6 horses

- *Sample space* $\Omega = \{\text{permutations of } (1,2,3,4,5,6)\}$
- Outcome $\omega = (i_1, i_2, i_3, i_4, i_5, i_6)$ has
  - ▷ horse i_1 in first place,
  - ▷ horse i_2 in second place, etc.
- The *event* "horse 2 wins the race" is the subset

$$A = \{ \text{ any permutation with 2 in first place } \} \subseteq \Omega,$$

- Another event:

$$B = \{\text{horse 3 comes in second place}\}.$$

- Then

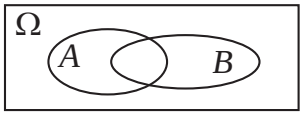
$$A \cap B = \{\text{horse 2 wins and horse 3 comes in second}\},$$
$$A \cup B = \{\text{horse 2 wins or horse 3 comes in second}\}.$$

- **Probability** $\mathcal{P}$ assigns to each event (set) $A \subseteq \Omega$ a number such that
  1. $0 \leq \mathcal{P}(A) \leq 1$ for any event $A \subseteq \Omega$.
  2. $\mathcal{P}(\Omega) = 1$.
  3. If $A_1, A_2, \ldots$ is (possibly infinite) sequence of disjoint events,

$$\mathcal{P}\left(\cup_i A_i\right) = \sum_i \mathcal{P}(A_i).$$

- Some properties of probability
  - $\mathcal{P}(\overline{A}) = 1 - \mathcal{P}(A)$
  - for all $A, B \subseteq \Omega$,

$$\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$$



Example cont. (flip fair coin three times)

- $\Omega = \{HHH,\ HHT, \ldots, TTT\}$
- $\mathcal{P}(\{\omega\}) = 1/8$ for each $\omega \in \Omega$.
- If $A = \{H$ on second toss and T on third toss$\}$, then
  - $A = \{HHT,\ THT\}$
  - $\mathcal{P}(A) = \sum_{\omega \in A} \mathcal{P}(\{\omega\}) = 1/4$
  - $\mathcal{P}(\overline{A}) = 1 - \mathcal{P}(A) = 3/4$
- If $A_1 = \{$ heads on first toss $\}$ and $A_2 = \{$ at least 2 tails $\}$,

$$\mathcal{P}(A_1 \cap A_2) = \mathcal{P}(\{HTT\}) = \frac{1}{8},$$

$$\begin{aligned}\mathcal{P}(A_1 \cup A_2) &= \mathcal{P}(A_1) + \mathcal{P}(A_2) - \mathcal{P}(A_1 \cap A_2) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{8} = \frac{7}{8}\end{aligned}$$

Example cont. (Race with 6 horses)

- Recall $A = \{$horse 2 wins the race$\}$.
- All race outcomes equally likely, i.e., $\mathcal{P}(\{\omega\}) = \frac{1}{6!}$, so

$$\mathcal{P}(A) = \sum_{\omega \in A} \mathcal{P}(\{\omega\}) = \frac{5!}{6!} = \frac{1}{6}$$

- If $B = \{$horse 3 comes in second place$\}$, then $\mathcal{P}(B) = 1/6$ and

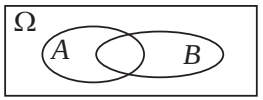
$$\mathcal{P}(A \cap B) = \frac{4!}{6!} = \frac{1}{6 \cdot 5} = \frac{1}{30},$$

$$\begin{aligned}\mathcal{P}(A \cup B) &= \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B) \\ &= \frac{1}{6} + \frac{1}{6} - \frac{1}{30} = \frac{3}{10}.\end{aligned}$$

Conditional Probability

Question : Given event A occurs, what is prob B occurs ?

Notation : $\mathcal{P}(B \mid A)$ is probability of B given A



- Suppose outcome $\omega \in \Omega$ occurs, and all we know is $\omega \in A$
- $\omega \in B$ if and only if $\omega \in BA$
- Thus,

$$\mathcal{P}(B \mid A) \;=\; \frac{\mathcal{P}(BA)}{\mathcal{P}(A)}$$

which implies

$$\begin{aligned}\mathcal{P}(BA) &= \mathcal{P}(B \mid A)\,\mathcal{P}(A) \\ &= \mathcal{P}(A \mid B)\,\mathcal{P}(B)\end{aligned}$$

Example:

- Roll a fair 6-sided die once.
- Event $A = \{\, 4 \text{ comes up } \}$
- Event $B = \{\, 3, 4, 5, \text{or } 6 \text{ comes up } \}$
- Then

$$\mathcal{P}(A \mid B) \;=\; \frac{\mathcal{P}(AB)}{\mathcal{P}(B)} \;=\; \frac{\mathcal{P}(A)}{\mathcal{P}(B)} \;=\; \frac{1/6}{4/6} \;=\; \frac{1}{4}$$

- Also

$$\mathcal{P}(B \mid A) \;=\; \frac{\mathcal{P}(AB)}{\mathcal{P}(A)} \;=\; \frac{\mathcal{P}(A)}{\mathcal{P}(A)} \;=\; 1$$

- What is $\mathcal{P}(A \mid \text{even number comes up})$?

Independent Events: Events A, B independent if

$$\mathcal{P}(A \mid B) = \mathcal{P}(A), \quad \mathcal{P}(B \mid A) = \mathcal{P}(B)$$

- A, B indep $\Rightarrow$ knowledge of occurrence of one event does not alter probability of other event
- From definition of conditional probability,

$$\begin{aligned}&\text{Events } A \text{ and } B \text{ independent} \\ \Leftrightarrow\;\; &\mathcal{P}(AB) = \mathcal{P}(A)\,\mathcal{P}(B)\end{aligned}$$

Example: flip fair coin three times

- Each outcome $\omega \in \Omega$ is a triple $\in \{H, T\}^3$
- For each $\omega \in \Omega$, $\mathcal{P}(\omega) = \frac{1}{8}$
- Are the following events A and B independent?
  - $A = \{\, H \text{ on first toss } \}$, $B = \{\, T \text{ on third toss } \}$
  - $A = \{\, H \text{ on first toss } \}$, $B = \{\, \text{at least one } T \,\}$

Random variables

- **Random variable** (RV) X is function from Ω to set $\mathbf{R}$ of real numbers, i.e., $X(\omega) \in \mathbf{R}$ for each outcome $\omega \in \Omega$
- Can translate probability $\mathcal{P}$ to RV X via relation

$$\mathcal{P}\big(X \in C\big) \;=\; \mathcal{P}\left(\big\{\omega \in \Omega : X(\omega) \in C\big\}\right)$$

for $C \subseteq \mathbf{R}$.

- **Example:** Flip coin three times:

$$\Omega = \{HHH,\; HHT,\; \ldots,\; TTT\}$$

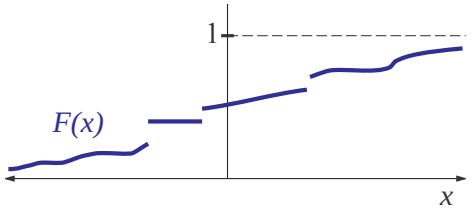
  - Define RV $X(\omega) = \#\ H$'s in outcome ω
  - If $\omega = HTH$, then $X(\omega) = 2$, and

$$\begin{aligned}\mathcal{P}(X = 2) &= \mathcal{P}(\{\omega \in \Omega : X(\omega) = 2\}) \\ &= \mathcal{P}(\{HHT,\; HTH,\; THH\}) \;=\; \frac{3}{8}\end{aligned}$$

- RV X defined by **cumulative distribution function (CDF)** F

$$F(x) \;=\; \mathcal{P}(X \le x), \quad x \in \mathbf{R}$$

- Properties of CDF:
  - $0 \le F(x) \le 1$
  - $F(x)$ is non-decreasing fcn
  - $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$



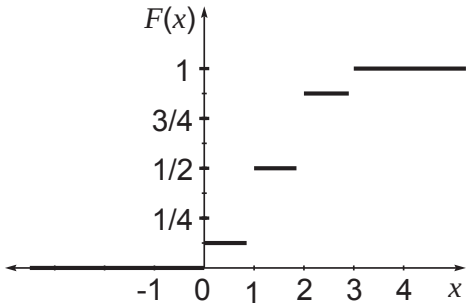
- **Notation:** $X \sim F$ means RV X has CDF or distribution F
- **Definition:** The **support** S of RV X is set of all possible values X can take on, so $\mathcal{P}(X \in S) = 1$.

Example: Flip fair coin three times

- Let X = number of heads, so

$$\mathcal{P}(X=0) \;=\; 1/8, \qquad \mathcal{P}(X=1) \;=\; 3/8$$
$$\mathcal{P}(X=2) \;=\; 3/8, \qquad \mathcal{P}(X=3) \;=\; 1/8$$

- Support: $\{0, 1, 2, 3\}$
- CDF $F(x) = \mathcal{P}(X \le x)$ for each $-\infty < x < \infty$:



Types of random variables :

1. Discrete :
   - Support is finite or countable set, e.g., $\mathcal{N} \equiv \{0, 1, 2, \ldots\}$ or $\{0, 1\}$.
   - Distn described by a **probability mass function** (PMF) $\mathbf{p}$ where $\mathbf{p}(x) = \mathcal{P}(X = x)$ and

$$\mathcal{P}(X \in A) \;=\; \sum_{x \in A} \mathbf{p}(x)$$

2. Continuous :
   - Support is a continuous set, e.g., $\mathbf{R}$ or $[a, b]$
   - Distn described by **probability density function** (PDF) $f(x)$ where $f(x) = \frac{dF(x)}{dx}$, i.e., $F(x) = \int_{-\infty}^{x} f(u)\, du$, and

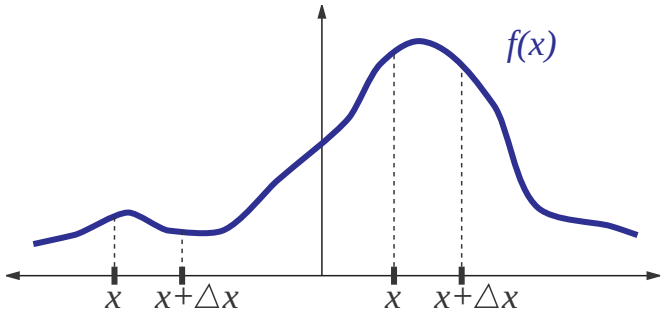
$$\mathcal{P}(X \in A) \;=\; \int_A f(x)\, dx$$

3. Mixed discrete and continuous (e.g., inches of rain tomorrow)

Interpretation of density:

- For any x and $\Delta x > 0$,

$$\mathcal{P}\{\, X \in [x,\, x + \Delta x]\,\} \;=\; \int_x^{x+\Delta x} f(u)\, du \;\approx\; f(x)\, \Delta x$$



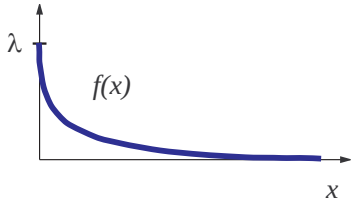
- X more likely to fall in region where density is high.

Example: Consider continuous RV X with

- Support $[0, \infty)$
- Density

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

where $\lambda > 0$ is a constant.



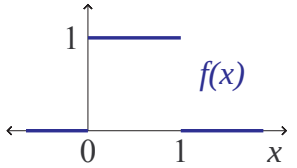
- CDF

$$F(x) = \mathcal{P}(X \leq x) = \int_{-\infty}^{x} f(y)\, dy = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

- Easy to show that F is indeed a CDF.
- $X \sim \text{Exp}(\lambda)$, exponential distribution with (rate) parameter λ

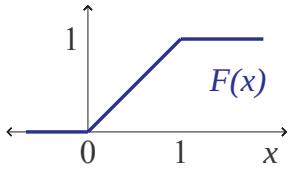
Example: $X \sim \text{unif}[0, 1]$ has density

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



and distribution function

$$F(x) = \begin{cases} 0 & \text{if } -\infty < x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } 1 < x < \infty \end{cases}.$$



Note that for $0 \leq x \leq x + \Delta x \leq 1$,

$$\begin{aligned} P\{\, X \in (x,\, x + \Delta x]\,\} &= F(x + \Delta x) - F(x) \\ &= (x + \Delta x) - x \;=\; \Delta x, \end{aligned}$$

which doesn't depend on x.

Multiple Random Variables

- **Random vector** $(X_1, X_2, \ldots, X_d)$ is collection of $d \geq 1$ random variables.
  - **Example:** For d stocks, let X_i be return of stock i, $i = 1, 2, \ldots, d$.
  - **Joint distribution** describes probabilistic joint behavior.
- To simplify things, let $d = 2$ and write $(X_1, X_2) = (X, Y)$.
- **Joint CDF** of (X, Y) is

$$F_{X,Y}(x, y) \;=\; \mathcal{P}\Big(\{X \leq x\} \cap \{Y \leq y\}\Big)$$

- **Marginal CDF** is CDF of one random variable by itself:

$$\begin{aligned} F_X(x) &= \mathcal{P}(X \leq x) = F_{X,Y}(x, \infty), \\ F_Y(y) &= \mathcal{P}(Y \leq y) = F_{X,Y}(\infty, y). \end{aligned}$$

- For (X, Y) jointly continuous,
  - **Joint density** is

$$f_{X,Y}(x, y) \;=\; \frac{\partial^2 F_{X,Y}(x, y)}{\partial x\, \partial y},$$

and for $A \subseteq \mathbf{R}^2$

$$\mathcal{P}\big((X, Y) \in A\big) \;=\; \int\!\!\int_A f_{X,Y}(x, y)\, dx\, dy$$

  - **Marginal densities** $f_X(x)$, $f_Y(y)$ are

$$\begin{aligned} f_X(x) &= \int_{\mathbf{R}} f_{X,Y}(x, y)\, dy \;=\; \frac{d}{dx} F_X(x), \\ f_Y(y) &= \int_{\mathbf{R}} f_{X,Y}(x, y)\, dx \;=\; \frac{d}{dy} F_Y(y). \end{aligned}$$

- For (X, Y) discrete RVs, can define joint and marginal PMFs.

- X, Y are **independent** if for all $C, D \subseteq \mathbf{R}$

$$\mathcal{P}\big((X,Y) \in C \times D\big) \;=\; \mathcal{P}(X \in C)\; \mathcal{P}(Y \in D)$$

- Jointly continuous random variables X, Y are **independent** if for all $x, y \in \mathbf{R}$, the joint and marginal densities satisfy

$$f_{X,Y}(x,y) \;=\; f_X(x)\; f_Y(y)$$

  - X, Y are **dependent** (not independent) if $\exists\; x$, y such that

$$f_{X,Y}(x,y) \;\neq\; f_X(x)\; f_Y(y)$$

- Discrete RVs X, Y are **independent** if joint PMF always equals product of marginal PMFs; otherwise, **dependent**.

Example: X, Y jointly discrete RVs with joint PMF

$$\mathbf{p}_{X,Y}(x,y) = \begin{cases} xy/27 & \text{for } x = 1,2,\; y = 2,3,4 \\ 0 & \text{otherwise} \end{cases}.$$

Then marginal PMFs are

$$\mathbf{p}_X(x) = \sum_y \mathbf{p}_{X,Y}(x,y) \;=\; \sum_{y=2}^{4} \frac{xy}{27} \;=\; \frac{x}{3} \quad \text{for } x = 1,2,$$

$$\mathbf{p}_Y(y) = \sum_x \mathbf{p}_{X,Y}(x,y) \;=\; \sum_{x=1}^{2} \frac{xy}{27} \;=\; \frac{y}{9} \quad \text{for } y = 2,3,4.$$

Thus,

$$\mathbf{p}_{X,Y}(x,y) \;=\; \frac{xy}{27} \;=\; \mathbf{p}_X(x)\; \mathbf{p}_Y(y) \quad \text{for all } x, y,$$

so X and Y are **independent**.

Example: X, Y jointly continuous RVs with joint density

$$f_{X,Y}(x,y) = \begin{cases} y - x & \text{for } 0 \le x \le 1 \text{ and } 1 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}.$$

- Marginal densities:

$$f_X(x) \;=\; \int_1^2 (y - x)\, dy \;=\; 3/2 - x \;\text{ for } 0 \le x \le 1.$$

$$f_Y(y) \;=\; \int_0^1 (y - x)\, dx \;=\; y - 1/2 \;\text{ for } 1 \le y \le 2.$$

- X and Y are **dependent** because

$$f_{X,Y}\,(0,\,1) \;=\; 1 \;\neq\; \left(\frac{3}{2}\right)\left(\frac{1}{2}\right) \;=\; f_X(0)\; f_Y(1)$$

Expectation, Variance, and Moments

- **Expectation** (or **mean**) of RV X is denoted by

$$\mu \;=\; \mathcal{E}[X]$$

  - if X is discrete with PMF $\mathbf{p}$ and support $\{x_1, \ldots, x_n\}$,

$$\mathcal{E}[X] \;=\; \sum_{i=1}^{n} x_i\, \mathbf{p}(x_i)$$

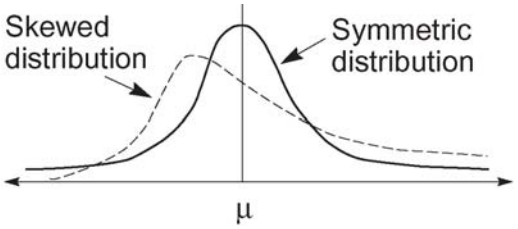
  - if X is continuous with density f,

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

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

Remarks on expectation:

- μ is weighted average of all possible values x of X
- For reasonably shaped distns, μ measures central tendency.
  - Mean often close to the **mode**, the single most likely value
  - For symmetric distributions, mean = **median**, which is the point x such that CDF $F(x) = 1/2$.



- **Example:** Flip fair coin three times, $X = \#$ heads

$$\mu = \sum_{i=1}^{n} x_i\, \mathrm{p}(x_i)$$
$$= 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

- **Example:** $X \sim$ unif$[0, 1]$

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases},$$

and so

$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 x \cdot 1\, dx = \frac{1}{2}.$$

- **Example:** $X \sim \mathrm{Exp}(\lambda)$, which is continuous RV with density

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

Then

$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^{\infty} x\, \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda}.$$

- Recall expectation:

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

- **Properties of expectation:**

1. $\mathcal{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$ for any function g.
2. $\mathcal{E}[aX + b] = a\, \mathcal{E}[X] + b$ for any constants $a, b \in \mathbf{R}$.
3. $\mathcal{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathcal{E}[X_i]$ (linearity, even if dependent).
4. If X, Y independent, then for any functions g and h,
$$\mathcal{E}[g(X)\, h(Y)] = \mathcal{E}[g(X)]\, \mathcal{E}[h(Y)].$$

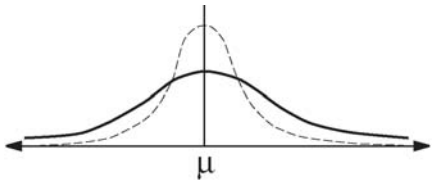
- kth **moment** of RV X:

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

- **Variance** $\mathcal{V}(X)$ of RV X with mean μ is

$$\mathcal{V}(X) \;=\; \mathcal{E}[(X-\mu)^2] \;=\; \int_{-\infty}^{\infty}(x-\mu)^2\, f(x)\, dx$$

- **Remarks on variance:**
  - Variance measures spread of distribution
    - ▷ How far from mean μ do typical values of X tend to fall
  - Variance of X often denoted σ^2 or σ_x^2
  - Can show $\sigma^2 = \mathcal{E}[X^2] - \mu^2$
  - **Standard deviation** $\sigma = \sqrt{\sigma^2}$
  - Variance in units^2, but standard deviation in same units as RV



- Recall variance $\mathcal{V}(X) \;=\; \mathcal{E}[(X-\mu)^2] \;=\; \mathcal{E}[X^2] - \mu^2$
- **Example:** Flip fair coin three times, $X = \#$ heads

$$\mathcal{E}[X^2] \;=\; 0^2\cdot\frac{1}{8} + 1^2\cdot\frac{3}{8} + 2^2\cdot\frac{3}{8} + 3^2\cdot\frac{1}{8} \;=\; 3,$$

and since $\mu = 3/2$, we have

$$\sigma^2 \;=\; 3 - \left(\frac{3}{2}\right)^2 \;=\; \frac{3}{4}.$$

- **Example:** $X \sim \text{Exp}(\lambda)$

$$\mathcal{E}[X^2] \;=\; \int_{-\infty}^{\infty} x^2\, f(x)\, dx \;=\; \int_0^{\infty} x^2\, \lambda e^{-\lambda x}\, dx \;=\; \frac{2}{\lambda^2},$$

and since $\mu = 1/\lambda$, we have

$$\sigma^2 \;=\; \mathcal{E}[X^2] - \mu^2 \;=\; \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \;=\; \frac{1}{\lambda^2}.$$

- Recall: variance $\mathcal{V}(X) \;=\; \mathcal{E}[(X-\mu)^2] \;=\; \mathcal{E}[X^2] - \mu^2$
- **Properties of variance:**
  1. $\mathcal{V}(X) \geq 0$ for any RV X
  2. $\mathcal{V}(X) = 0$ iff $X =$ constant with probability 1
  3. for all constants $a, b \in \mathbf{R}$,

$$\mathcal{V}[aX + b] = a^2\mathcal{V}[X]$$

- kth **central moment** of RV X:

$$\alpha_k \;=\; \mathcal{E}[(X-\mu)^k] \;=\; \int_{-\infty}^{\infty}(x-\mu)^k\, f(x)\, dx$$

Covariance

- **Covariance** of RVs X, Y with $\mu_x = \mathcal{E}[X]$, $\mu_y = \mathcal{E}[Y]$:

$$\text{Cov}(X,Y) \;=\; \mathcal{E}[(X-\mu_x)(Y-\mu_y)] \;=\; \mathcal{E}[XY] - \mu_x\,\mu_y$$

- **Properties of covariance:**
  1. If X, Y independent, then $\text{Cov}(X,Y) = 0$
     - Converse holds (i.e., $\text{Cov}(X,Y) = 0 \;\Rightarrow\; X, Y$ indep) if X and Y are normally distributed, but not in general.
  2. Variance of $X + Y$:

$$\mathcal{V}(X+Y) \;=\; \mathcal{V}(X) + \mathcal{V}(Y) + 2\,\text{Cov}(X,Y)$$

  3. If X, Y independent, then $\mathcal{V}(X+Y) = \mathcal{V}(X) + \mathcal{V}(Y)$.
- **Definition:** X, Y **positively correlated** if $\text{Cov}(X,Y) > 0$.
- **Definition:** X, Y **negatively correlated** if $\text{Cov}(X,Y) < 0$.

Remarks:

- Recall $\text{Cov}(X,Y) = \mathcal{E}[(X-\mu_x)(Y-\mu_y)]$
- If X, Y are positively correlated, then
  - $X > \mu_x$ and $Y > \mu_y$ tend to occur together;
  - $X < \mu_x$ and $Y < \mu_y$ tend to occur together.
- If X, Y are negatively correlated, then
  - $X > \mu_x$ and $Y < \mu_y$ tend to occur together;
  - $X < \mu_x$ and $Y > \mu_y$ tend to occur together.
- Covariance measures **linear dependence**, but depends on units.
- To get unitless measure of dependence, use **correlation**

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_x\,\sigma_y}$$

- Can show that $-1 \le \rho(X,Y) \le 1$.

Example: X, Y jointly continuous RVs with joint density

$$f_{X,Y}(x,y) = \begin{cases} y-x & \text{for } 0 \le x < 1 \text{ and } 1 \le y < 2 \\ 0 & \text{otherwise} \end{cases}.$$

- Marginal densities (from p. 2-23)

$$\begin{aligned} f_X(x) &= 3/2 - x \quad \text{for} \quad 0 \le x \le 1, \\ f_Y(y) &= y - 1/2 \quad \text{for} \quad 1 \le y \le 2. \end{aligned}$$

- Moments of X:

$$\begin{aligned} \mu_x &= \int_{-\infty}^{\infty} x\, f_X(x)\,dx = \int_0^1 x(3/2-x)\,dx = 5/12, \\ \mathcal{E}(X^2) &= \int_{-\infty}^{\infty} x^2\, f_X(x)\,dx = \int_0^1 x^2(3/2-x)\,dx = 1/4, \\ \sigma_x^2 &= \mathcal{E}(X^2) - \mu_x^2 = 11/144 \end{aligned}$$

- Moments of Y:

$$\begin{aligned} \mu_y &= \int_{-\infty}^{\infty} y\, f_Y(y)\,dy = \int_1^2 y(y-1/2)\,dy = 19/12, \\ \mathcal{E}(Y^2) &= \int_{-\infty}^{\infty} y^2\, f_Y(y)\,dy = \int_1^2 y^2(y-1/2)\,dy = 31/12, \\ \sigma_y^2 &= \mathcal{E}(Y^2) - \mu_y^2 = 21/144 \end{aligned}$$

- Compute covariance and correlation of X and Y:

$$\mathcal{E}(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{X,Y}(x,y)\,dx\,dy = \int_0^1\int_1^2 xy(y-x)\,dy\,dx = \frac{2}{3}.$$

Therefore,

$$\text{Cov}(X,Y) = \mathcal{E}(XY) - \mu_x\,\mu_y = \frac{1}{144}$$

and

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_x\,\sigma_y} = \frac{1/144}{\sqrt{(11/144)(21/144)}} \approx 0.0658$$

Sample Moments

- Observe a **sample** of n observations from a distribution F
  - $X_1, X_2, \ldots, X_n$ are **independent and identically distributed (IID)** random variables from distn F.
- An estimate of mean μ of distn F is the **sample mean**:

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

  - In Excel, use `AVERAGE` function
- An estimate of variance σ^2 of distn F is the **sample variance**:

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}_n\right)^2 = \left(\frac{1}{n-1}\sum_{k=1}^{n} X_k^2\right) - \frac{n}{n-1}\bar{X}_n^2$$

  - In Excel, use `VAR` function
- What is an estimator for standard deviation σ?

Important:

- Sample mean $\bar{X}_n$ and sample variance S_n^2 are **random** since they depend on random sample $X_1, X_2, \ldots, X_n$
- True (population) mean μ and true variance σ^2 are constants
  - Sample mean $\bar{X}_n$ is not the same as true mean μ
  - Sample variance S_n^2 is not the same as true variance σ^2
- When $X_1, X_2, \ldots, X_n$ are IID,

$$\begin{aligned}\mathcal{V}(\bar{X}_n) &= \mathcal{V}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\mathcal{V}\left(\sum_{k=1}^{n} X_k\right) \\ &= \frac{1}{n^2}\sum_{k=1}^{n}\mathcal{V}(X_k) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.\end{aligned}$$

Specific Parametric Families of Continuous Distributions

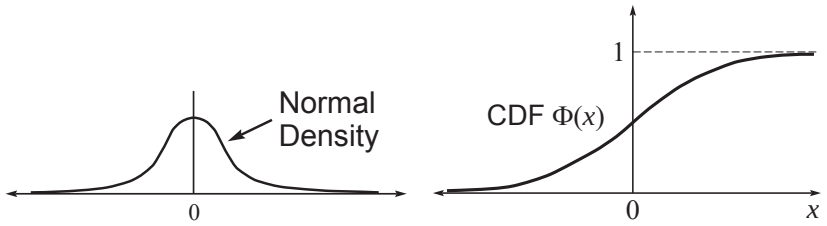
- Support is continuous set
- Distn defined by **probability density function** (PDF) f
- Examples
  - Normal
  - Chi-square
  - Student t distribution
  - Many other distributions; e.g., see N. L. Johnson, S. Kotz and N. Balakrishnan. 1994. *Continuous Univariate Distributions*, 2nd edition. Wiley-Interscience; ISBN: 0471584959.

Normal (Gaussian)

- Density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty,$$

  where
  - μ = mean
  - σ = standard deviation
- **Notation:**
  - $\mathcal{N}(\mu, \sigma^2)$ is normal distn with mean μ and variance σ^2.
  - $\mathcal{N}(0, 1)$ is called the **standard normal** distn
  - $\Phi(x)$ denotes CDF of $\mathcal{N}(0, 1)$



- CDF of $\mathcal{N}(\mu, \sigma^2)$ cannot be written in closed form
  - Can look up values of CDF in z-table after suitable transformation
  - Excel functions provide numerical approximations
    - ▷ `NORMSDIST` (or `NORMDIST`) for CDF
    - ▷ `NORMSINV` (or `NORMINV`) for inverse of CDF

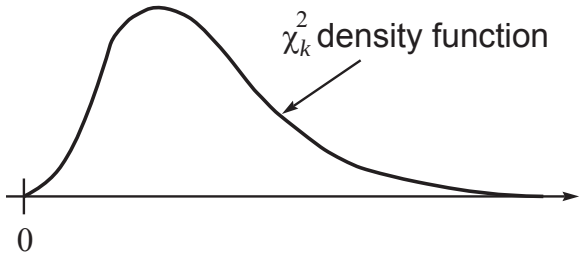
- Suppose that quantity X that we are trying to model can be expressed as sum of random variables:
$$X \;=\; Y_1 + Y_2 + \cdots + Y_n.$$
- In great generality, **central limit theorem** asserts that, for n large,
$$X \;\overset{\mathcal{D}}{\approx}\; \mathcal{N}(\mu_n, \sigma_n^2)$$
for some constants μ_n and σ_n^2, where $\overset{\mathcal{D}}{\approx}$ means "approximately equal in distribution to."
- Y_i's need not be identically distributed nor independent.
- **Moral of Story:** If X is sum of large number of random quantities, then X often can be approximately modeled as normal random variable.

Properties of Normal Distribution:

- Support: $(-\infty, \infty)$
- If $X \sim \mathcal{N}(\mu,\ \sigma^2)$, then $Z \;=\; (X-\mu)/\sigma \;\sim\; \mathcal{N}(0,1)$.
- If $X_1, X_2, \ldots, X_n$ are indep normals with $X_i \sim \mathcal{N}(\mu_i,\ \sigma_i^2)$, then
$$\sum_{i=1}^n X_i \;\sim\; \mathcal{N}\left(\sum_{i=1}^n \mu_i,\ \sum_{i=1}^n \sigma_i^2\right).$$
- If $X_1, X_2, \ldots, X_n$ are IID $\mathcal{N}(\mu, \sigma^2)$, then
  - $\bar{X}_n$ and S_n^2 are independent
  - $\bar{X}_n \sim \mathcal{N}(\mu,\ \sigma^2/n)$
  - Expect approximately 68% (resp., 95%, 99.7%) of observations $X_1, X_2, \ldots, X_n$ to lie within 1 (resp., 2, 3) standard deviations σ of mean μ.

Chi-Square Distribution

- Suppose $Z_1, Z_2, \ldots, Z_k$ are IID $\mathcal{N}(0,1)$
- Then
$$Y \;=\; Z_1^2 + Z_2^2 + \cdots + Z_k^2 \;\sim\; \chi_k^2,$$
which is χ^2 (chi-square) distn with k degrees of freedom (df)
- Density function of χ_k^2 :
$$f_k(x) \;=\; \frac{1}{2^{k/2}\Gamma(k/2)}\, x^{(k/2)-1}\, e^{-x/2}\, I_{[0,\infty)}(x)$$
where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$ and for any set A,
$$I_A(x) \;=\; \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$
- Support: $[0, \infty)$
- Mean $= k$, Variance $= 2k$



- CDF of χ^2 cannot be evaluated in closed form.
  - See course webpage for table giving critical values $\chi^2_{\alpha,k}$ for χ_k^2; i.e., $P(\chi_k^2 > \chi^2_{\alpha,k}) = \alpha$
  - In Excel, use function
    - ▷ `CHIDIST` for tail CDF
    - ▷ `CHIINV` for inverse of CDF

Remarks on Chi-Square Distribution

- If $Z_1, Z_2, \ldots, Z_n \sim \mathcal{N}(0,1)$ IID and $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$, then $\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 \sim \chi^2_{n-1}$
- If $X_1, X_2, \ldots, X_n$ are IID $\mathcal{N}(\mu, \sigma^2)$, then the following have χ^2_{n-1} distributions:
  - $\sum_{i=1}^n (X_i - \bar{X}_n)^2/\sigma^2$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^k X_i$
  - $(n-1)\, S_n^2 \,/\, \sigma^2$, where $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

  Key Points:
  - If sampling IID normals, then $(n-1)S_n^2/\sigma^2$ has χ^2_{n-1} distn.
  - Allows one to draw exact inferences about variance σ^2.
  - Generally speaking, sums of squares of IID normals can be massaged into χ^2

Student t Distribution

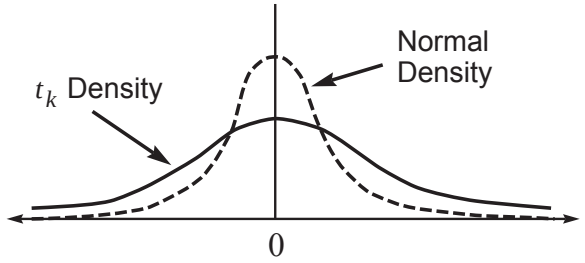
- t random variable with k df, $k > 0$:

$$t_k = \frac{Z}{\sqrt{U/k}}$$

  - $Z \sim \mathcal{N}(0,1),\ U \sim \chi^2_k$
  - Z and U independent
- Density:

$$f(x) = \left(\frac{\Gamma[(k+1)/2]}{\Gamma(k/2)}\right)\left(\frac{1}{\sqrt{k\pi}}\right)\left(\frac{1}{(1+x^2/k)^{(k+1)/2}}\right)$$

- Support: $(-\infty, \infty)$
- Mean: 0, Variance: $\frac{k}{k-2}$ when $k > 2$



- CDF of t_k cannot be evaluated in closed form.
  - See course webpage for table giving critical values $\tau_{\alpha,k}$ for t_k; i.e., $P(t_k > \tau_{\alpha,k}) = \alpha$
  - In Excel, use function
    - ▷ `TDIST` for CDF
    - ▷ `TINV` for critical values of *2 tails*

Properties of t Distribution

- t_k distn has fatter tails than normal.
- As df $k \to \infty$,

$$\mathcal{P}(t_k \le x) \ \to\ \mathcal{P}(\mathcal{N}(0,1) \le x) \ \text{ for each } x \in \mathbf{R}$$

- Let $X_1, X_2, \ldots, X_n$ be IID $\mathcal{N}(\mu, \sigma^2)$. Then RV

$$T \ = \ \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}}$$

  has t distribution with $n-1$ df.

- **Key Points:**
  - If sampling n IID normals, then T has t_{n-1} distribution.
  - Allows one to draw exact inferences about mean μ when variance is also unknown.

Summary of Distributions

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim \mathcal{N}(0, 1)$.
- If $Z_1, \ldots, Z_k \sim \mathcal{N}(0, 1)$ IID, then $Z_1^2 + \cdots + Z_k^2 \sim \chi_k^2$
- If $Z \sim \mathcal{N}(0, 1)$ and indep $U \sim \chi_k^2$, then $T = \frac{Z}{\sqrt{U/k}} \sim t_k$
- If $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ IID, then

$$(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2 \quad \text{and} \quad \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \sim t_{n-1}$$

- Sums of squares of IID normals can be massaged into χ^2
- Many relationships exist among different distributions

Point Estimators

- **Goal:** estimate parameter θ of or derived from a distn F, e.g.,
  - Moments $\mathcal{E}[X^k]$ or central moments $\mathcal{E}[(X - \mu)^k]$
  - Parameter λ of exponential distn
  - 99% value-at-risk ξ; i.e., $F(\xi) = 0.99$
- **General Approach:**
  - Draw sample of n IID observations $X_1, X_2, \ldots, X_n$ from F.
  - Compute **point estimator** $\hat{\theta}_n$ of θ based on the sample.

- Examples of point estimators:
  - An estimator of mean μ is the **sample mean**

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

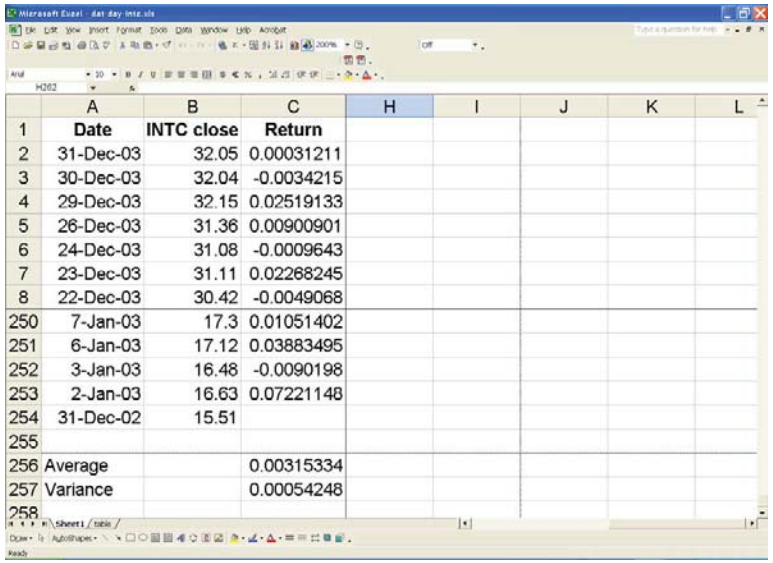
  - An estimator of variance σ^2 is the **sample variance**

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2$$

  - An estimator of covariance $\mathrm{Cov}(X, Y) = \mathcal{E}[(X - \mu_x)(Y - \mu_y)]$ is **sample covariance**

$$\widehat{\mathrm{Cov}}_n = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$$

Example: Intel (INTC) Daily Closing Prices and Returns



	A	B	C	H	I	J	K	L
1	Date	INTC close	Return					
2	31-Dec-03	32.05	0.00031211					
3	30-Dec-03	32.04	-0.0034215					
4	29-Dec-03	32.15	0.02519133					
5	26-Dec-03	31.36	0.00900901					
6	24-Dec-03	31.08	-0.0009643					
7	23-Dec-03	31.11	0.02268245					
8	22-Dec-03	30.42	-0.0049068					
250	7-Jan-03	17.3	0.01051402					
251	6-Jan-03	17.12	0.03883495					
252	3-Jan-03	16.48	-0.0090198					
253	2-Jan-03	16.63	0.07221148					
254	31-Dec-02	15.51						
255								
256	Average		0.00315334					
257	Variance		0.00054248					
258								

- Daily closing prices for INTC for 2003
- Let P_i = closing price for INTC on day i
- Return on ith day:

$$X_i = \frac{P_i - P_{i-1}}{P_{i-1}} = \frac{P_i}{P_{i-1}} - 1$$

- $n = 252$ returns
- Then
  - Estimator of expected daily return is

$$\bar{X}_n \;=\; 0.00315$$

  - Estimator of variance of daily return is

$$S_n^2 \;=\; 0.00054$$

Assessing Quality of Point Estimators

- Suppose that $\widehat{\theta}_n$ is an estimator of parameter θ based on sample of n IID observations $X_1, \ldots, X_n$.
- **Definition:** **Standard error** of $\widehat{\theta}_n$ is

$$\mathcal{SE}(\widehat{\theta}_n) \;=\; \sqrt{\mathcal{V}(\widehat{\theta}_n)}$$

- **Example:** Sample mean $\bar{X}_n$
  - $\mathcal{SE}(\bar{X}_n) \;=\; \sqrt{\mathcal{V}(\bar{X}_n)} \;=\; \sigma/\sqrt{n}$ from before (p. 2-37)
  - **Estimated** standard error of $\bar{X}_n$ is $S_n/\sqrt{n}$
- **Example:** Sample variance S_n^2 of $\mathcal{N}(\mu, \sigma^2)$ data
  - Can show that $\mathcal{V}(S_n^2) = \frac{2\sigma^4}{n-1}$
  - $\mathcal{SE}(S_n^2) = \sqrt{\frac{2\sigma^4}{n-1}}$

- **Definition:** $\mathcal{E}[\widehat{\theta}_n] - \theta$ is the **bias** of estimator $\widehat{\theta}_n$
  - If bias $= 0$, i.e., $\mathcal{E}[\widehat{\theta}_n] = \theta$, then estimator is **unbiased**.
  - If $\mathcal{E}[\widehat{\theta}_n] \neq \theta$, then $\widehat{\theta}_n$ is a **biased** estimator of θ.
- **Example:** Suppose that $\bar{X}_n$ is sample mean of IID $X_1, X_2, \ldots, X_n$ having true mean μ. Then

$$\begin{aligned}\mathcal{E}[\bar{X}_n] &= \mathcal{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \;=\; \frac{1}{n}\mathcal{E}\left[\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}\left[X_i\right] \;=\; \frac{1}{n}(n\mu) \;=\; \mu\end{aligned}$$

  so $\bar{X}_n$ is unbiased estimator for μ.
- Can also show S_n^2 is unbiased estimator for σ^2.
- **Definition:** **Mean-squared error** of estimator $\widehat{\theta}_n$ of θ:

$$\mathcal{MSE}(\widehat{\theta}_n) \;=\; \mathcal{E}[(\widehat{\theta}_n - \theta)^2] \;=\; [\text{Bias}(\widehat{\theta}_n)]^2 + \mathcal{V}(\widehat{\theta}_n)$$

Summary of Point Estimators

- Often interested in parameter θ of unknown distn F, e.g.,
  - Mean μ and variance σ^2
- **Approach:**
  - Draw sample of n IID observations $X_1, X_2, \ldots, X_n$ from F
  - From sample, compute (point) estimator $\widehat{\theta}_n$ of θ
- Estimator $\widehat{\theta}_n$ is *not* same as true parameter value θ.
- Estimator $\widehat{\theta}_n$ has error.
  - Assess error in estimator $\widehat{\theta}_n$ by bias, standard error, $\mathcal{MSE}$, ...
  - All estimators $\widehat{\theta}_n$ we will consider are **consistent**, i.e., $\widehat{\theta}_n \approx \theta$ for large n.

Law of Large Numbers

- Suppose $X_1, X_2, \ldots, X_n$ is sequence of IID random variables with mean $\mu = \mathcal{E}[X]$ and variance $\sigma^2 = \mathcal{V}[X]$.
- Estimator of the true mean μ is sample mean

$$\bar{X}_n \;=\; \frac{1}{n}\sum_{k=1}^{n} X_k$$

- **Law of large numbers (LLN)**: If $\mathcal{E}[|X|] < \infty$, then

$$\bar{X}_n \;\approx\; \mu \quad \text{for large } n$$

  - Strong law of large numbers:

$$\mathcal{P}\left\{\lim_{n\to\infty} \bar{X}_n = \mu\right\} \;=\; 1.$$

  - Weak law of large numbers: for every $\delta > 0$,

$$\lim_{n\to\infty} \mathcal{P}\left\{|\bar{X}_n - \mu| < \delta\right\} \;=\; 1.$$

- But how close are $\bar{X}_n$ and μ ?

Central Limit Theorems

- **Central Limit Theorem (CLT)**: If $0 < \sigma^2 < \infty$, then

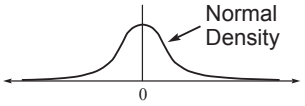
$$\frac{\sqrt{n}}{\sigma}\left[\bar{X}_n - \mu\right] \;\overset{\mathcal{D}}{\approx}\; \mathcal{N}(0,1) \quad \text{for large } n$$

  where $\overset{\mathcal{D}}{\approx}$ means "has approximately same distribution as"

  - More precisely, for every $x \in \mathbf{R}$,

$$\lim_{n\to\infty} \mathcal{P}\left\{\frac{\sqrt{n}}{\sigma}\left[\bar{X}_n - \mu\right] \le x\right\} \;=\; \mathcal{P}\left\{\mathcal{N}(0,1) \le x\right\}$$

- **Rules of Thumb:**
  - CLT approximately holds for $n \ge 30$
  - If we replace σ by its estimator S_n, then CLT needs $n \ge 50$



- **Interpretation of CLT :** Recall

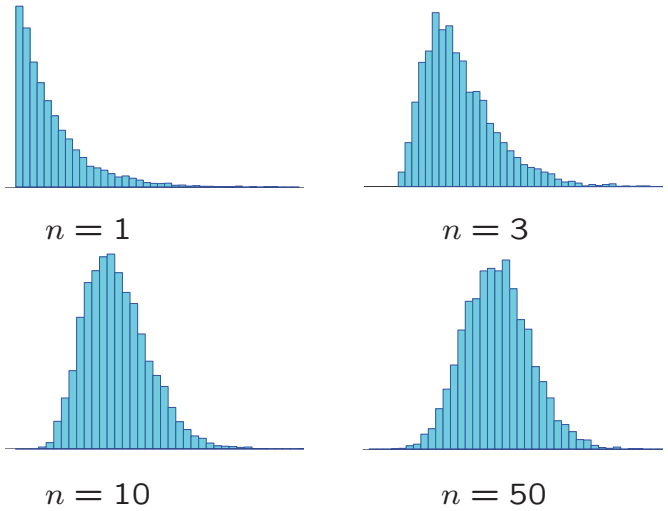
$$\frac{\sqrt{n}}{\sigma}\left[\,\bar{X}_n - \mu\right] \;\overset{\mathcal{D}}{\approx}\; \mathcal{N}(0,1)$$

- After a little algebra,

$$\bar{X}_n \;\overset{\mathcal{D}}{\approx}\; \mu + \frac{\sigma}{\sqrt{n}}\mathcal{N}(0,1) \;=\; \mathcal{N}(\mu, \sigma^2/n)$$

- *Error* in $\bar{X}_n$
  - Depends on sample size n through $1/\sqrt{n}$.
  - Depends on sampling distn F through σ.
  - Normally distributed.
- Also, $\sum_{i=1}^{n} X_i \;=\; n\bar{X}_n \;\overset{\mathcal{D}}{\approx}\; \mathcal{N}(n\mu, n\sigma^2)$

- Experiment: sample of n IID exponentials $X_1, X_2, \ldots, X_n$
- Histograms of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ (500 experiments)



Confidence Intervals

- **Goal:** provide measure of error of point estimator $\widehat{\theta}_n$ of parameter θ of distn F
- **Approach:** Provide error bounds for θ with "high confidence"
- **Example:** INTC data

$$\bar{X}_n = 0.00315$$

$$S_n^2 = 0.00054$$

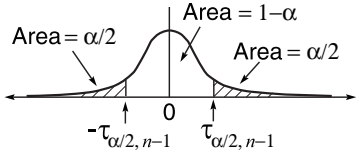
  - How close is point estimate $\bar{X}_n$ to true value μ ?
  - How close is point estimate S_n^2 to true value σ^2 ?

Confidence Intervals for Mean of Normal

- Sample of n IID observations $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, where μ, σ^2 unknown.
- Compute sample mean $\bar{X}_n$ and sample variance S_n^2, and know

$$\frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \sim t_{n-1}$$

- Fix **confidence level** $0 < 1 - \alpha < 1$, e.g., $\alpha = 0.05$ or 0.01, so $1 - \alpha = 0.95$ or 0.99
- In t-table, look up value of constant $\tau \equiv \tau_{\alpha/2, n-1}$ so $\mathcal{P}(t_{n-1} \geq \tau) = \alpha/2$.



- Then

$$\begin{aligned} 1 - \alpha &= \mathcal{P}(-\tau \leq t_{n-1} \leq \tau) \\ &= \mathcal{P}\left(-\tau \leq \frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \leq \tau \right) \\ &= \mathcal{P}\left(-\tau \frac{S_n}{\sqrt{n}} \leq \bar{X}_n - \mu \leq \tau \frac{S_n}{\sqrt{n}} \right) \\ &= \mathcal{P}\left(\bar{X}_n - \tau \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \tau \frac{S_n}{\sqrt{n}} \right) \end{aligned}$$

- Thus, we say that the interval

$$I_n \equiv \left[\bar{X}_n \pm \tau \frac{S_n}{\sqrt{n}} \right]$$

is a $100(1-\alpha)\%$ **confidence interval (CI)** for μ.

Procedure to Construct CI for Mean μ of Normal:

1. Collect sample of $n \geq 2$ IID observations $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$
2. Calculate sample mean $\bar{X}_n$ and sample variance S_n^2:

$$\begin{aligned} \bar{X}_n &= \frac{1}{n} \sum_{i=1}^{n} X_i, \\ S_n^2 &= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2 \end{aligned}$$

3. *Exact* $100(1-\alpha)\%$ CI for μ:

$$I \equiv \left[\bar{X}_n \pm \tau \frac{S_n}{\sqrt{n}} \right]$$

where $\tau = \tau_{\alpha/2,\, n-1}$ is the $\alpha/2$ critical point of a t-distn with $n-1$ df. Interval I is known as a t-CI.

- **95% CI for mean return μ of INTC**

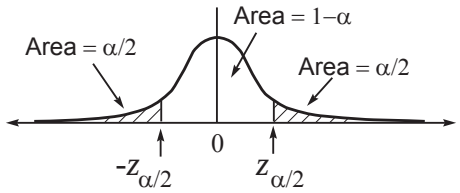
$$I = (0.00026,\ 0.00604)$$

- **Interpretation of $100(1-\alpha)\%$ Confidence Interval:**
  - **Correct:** $100(1-\alpha)\%$ confident that true mean $\mu \in I$
  - **Incorrect:** Future single observation has probability $1-\alpha$ of being in I. [For this, you need **prediction interval** (PI).]
  - **Incorrect:** $100(1-\alpha)\%$ of mass of distn lies in I. [For this, you need **tolerance interval**.]
- If $X_1, \ldots, X_n$ not normally distributed, can still construct approx CI for mean μ by using CLT
  - but now require large sample sizes $n \geq 50$.

- To construct an approx. $100(1-\alpha)\%$ CI for μ based on CLT
  - Look up in z-table value of constant $z \equiv z_{\alpha/2}$ such that

$$P(\mathcal{N}(0,1) \geq z) \ = \ \alpha/2,$$

e.g., if $\alpha = 0.05$, then $z = 1.96$.



  - For large sample sizes n, CLT ensures

$$\frac{\sqrt{n}}{S_n}(\bar{X}_n - \mu) \ \stackrel{\mathcal{D}}{\approx} \ \mathcal{N}(0,1)$$

- Therefore,

$$\begin{aligned} 1-\alpha \ &= \ \mathcal{P}\,(\ -z \ \leq \ \mathcal{N}(0,1) \ \leq \ z\ ) \\ &\approx \ \mathcal{P}\left(\ -z \ \leq \ \frac{\sqrt{n}}{S_n}(\bar{X}_n - \mu) \ \leq \ z\ \right) \\ &= \ \mathcal{P}\left(\ -z\frac{S_n}{\sqrt{n}} \ \leq \ \bar{X}_n - \mu \ \leq \ z\frac{S_n}{\sqrt{n}}\ \right) \\ &= \ \mathcal{P}\left(\ \bar{X}_n - z\frac{S_n}{\sqrt{n}} \ \leq \ \mu \ \leq \ \bar{X}_n + z\frac{S_n}{\sqrt{n}}\ \right) \end{aligned}$$

- Approximate $100(1-\alpha)\%$ CI for mean μ is

$$I_n \ \equiv \ \left[\ \bar{X}_n \ \pm \ z\frac{S_n}{\sqrt{n}}\ \right]$$

Remarks:

- CI is random since its endpoints depend on the sample mean $\bar{X}_n$ and sample variance S_n^2, which are random.
- Thus, if we generate many different CIs, then some will contain the true mean μ and some will not.
- We expect that for n large, approximately $100(1-\alpha)\%$ of the CIs will contain μ.
- The actual probability that the CI contains μ is called the **coverage** of the CI.
- When $X_1, X_2, \ldots, X_n$ are IID normal, coverage of t-CI is exact:

$$\mathcal{P}\left\{\mu \ \in \ \left[\ \bar{X}_n \ \pm \ \tau_{\alpha/2,n-1}\frac{S_n}{\sqrt{n}}\ \right]\right\} \ = \ 1-\alpha$$

for any sample size $n \geq 2$.

- When $X_1, X_2, \ldots, X_n$ are IID but **not normal**, CI based on CLT does not have exactly correct coverage:

$$\mathcal{P}\left\{ \mu \in \left[\bar{X}_n \pm z_{\alpha/2}\frac{S_n}{\sqrt{n}} \right] \right\} \approx 1 - \alpha$$

  - Need sample size $n \geq 50$ for approximation to roughly hold.
  - Coverage of CI often less than $100(1-\alpha)\%$.

- In general, to make width of CI decrease by a factor of $1/2$,
  - we must increase the sample size n by a factor of 4.

Problems Arising From Dependence

- All of the previous results from classical statistics assumes that $X_1, X_2, \ldots, X_n$ are IID.
- However, this is sometimes not the case in simulation
  - e.g., consider the successive waiting times of customers in a queue, which are dependent.
- In general, one often considers a so-called (discrete-time) *stochastic process* $[X_i : i = 1, 2, 3, \ldots]$
  - Subscript i denotes what happens at time i
  - e.g., X_i is the waiting time of the ith customer.
  - X_i and X_j may be dependent for all i, j.

- Can also consider a continuous-time stochastic process $[X(t) : t \geq 0]$
  - $X(t)$ represents the state of the system at time t
  - $X(t)$ may be number of customers in system at time t.

- Dependence of RVs encountered in simulation can cause problems if we try to apply classical statistics directly.
  - Must use other techniques to analyze simulation output.

Summary of Confidence Intervals

- CI provides probabilistic error bounds for unknown true mean μ.
- $100(1-\alpha)\%$ confident that μ lies in CI
- If normal data, construct CI using t-intervals, which is **exact** for any sample size $n \geq 2$.
- If non-normal data, construct CI using CLT approximation but need large samples ($n \geq 50$) to be approximately valid.
- Can also construct CI for other parameters, e.g., σ^2.

Stochastic Processes

- A **stochastic process** is a collection of random variables $[X_t : t \in I]$ for some index set I

Examples:

1. Outcome of coin tosses: $X_t = 1$ if t-th toss is heads; else 0.
   - for each t, $X_t \in \{0, 1\}$ (discrete)
   - the index set $I = \mathbf{N}$, the set of natural numbers (discrete)
2. Temperature from 1:00pm to 2:00pm tomorrow
   - for each t, $X_t \in \mathbf{R}$ (continuous)
   - index set $I =$ [1:00pm, 2:00pm] $\subset \mathbf{R}$ (continuous)
3. Sequence of arrival times $[X_t : t = 1, 2, \ldots]$ of customers
   - for each t, $X_t \geq 0$ (continuous)
   - index set $I = \mathbf{N}$, the set of natural numbers (discrete)

- Special cases of interest to us :

1. **Markov Process**: A stochastic process $[X_t : t \in I]$ is *Markov* if for all $t, t + h \in I$ and $A \subseteq \mathbf{R}$,

$$\mathcal{P}\Big(X_{t+h} \in A \mid X_s \text{ for all } s \leq t\Big) \;=\; \mathcal{P}\Big(X_{t+h} \in A \mid X_t\Big)$$

   - Given the present X_t, the future X_{t+h} is independent of the past X_s, $s < t$.
   - Do not have to keep track of the past for simulation.
   - Defining the proper *state* helps convert process into a Markov process.

Example: Number $N(t)$ of cars served by time t if we assume exponential inter-arrival and service times.

2. **Poisson process**: A special Markov process
   - Models *events* that are happening randomly in time.
   - Let $N(t)$ is the number of events that occur by time t
   - $[N(t) : t \geq 0]$ is called a *Poisson process with rate* λ if
     - ▷ $N(0) = 0$
     - ▷ Inter-event times are independent.
     - ▷ Time between events have exponential distn with mean $1/\lambda$.
   - Using these properties, can show that

$$\mathcal{P}(N(t) = k) \;=\; e^{-\lambda t}\frac{(\lambda t)^k}{k!},$$

     so $N(t)$ has Poisson distribution with parameter λt, and $\mathcal{E}[N(t)] \;=\; \mathcal{V}[N(t)] \;=\; \lambda t$.
   - *A Poisson process does not have Poisson-distributed inter-event times!*



- Let $[X_i : i \geq 1]$ be the sequence of inter-event times of a Poisson process, i.e.,
  - ▷ X_1 is the time when the first event happens
  - ▷ X_n is the time elapsed between the $(n-1)$th and nth event, $n \geq 2$
- The sequence $[X_i : i \geq 1]$ is IID.
- Moreover, X_j has an exponential distribution with mean $\frac{1}{\lambda}$, i.e.,

$$\mathcal{P}(X_j \leq t) \;=\; \begin{cases} 1 - e^{-\lambda t} & t \geq 0, \\ 0 & t < 0 \end{cases}$$

- An important property of the exponential distribution is the **memoryless property** ...
  - ▷ If $X \sim \exp(\lambda)$ and $t \geq s$, then

$$\begin{aligned}\mathcal{P}(X \geq t \mid X \geq s) &= \frac{\mathcal{P}(X \geq t,\, X \geq s)}{\mathcal{P}(X \geq s)} \\ &= \frac{\mathcal{P}(X \geq t)}{\mathcal{P}(X \geq s)} = \frac{e^{-\lambda t}}{e^{-\lambda s}} \\ &= e^{-\lambda(t-s)} = \mathcal{P}(X \geq t-s)\end{aligned}$$

  - ▷ Intuition: The past of the variable (memory) has no information about on how large it will be.

Hypothesis Testing

Example: INTC statistics

$$\bar{X}_n = 0.00315$$

$$S_n^2 = 0.00054$$

- Is INTC's true mean return $\mu = 0$?
- **Goal:** to answer question in statistically meaningful way.

Approach: hypothesis testing

- Hypothesis-testing paradigm calls for selection of
  - **null hypothesis** H_0
  - **alternative hypothesis** H_a
- Compute **test statistic** T, which is some function of our sample and whose distribution is known when H_0 is true.
- If value of test statistic T is unlikely to occur when H_0 is true, then reject H_0. Specifically,
  - Specify **(significance) level** α of test, e.g., 0.05 or 0.01.
  - Define **critical region** C. If H_0 true, then $\mathcal{P}(T \in C) = \alpha$.
  - If test statistic T falls in C, then H_0 is **rejected**; otherwise, H_0 is **not rejected** (or "accepted").
- Hypothesis test gives H_0 benefit of the doubt.

Example: Assume that $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ IID with unknown mean μ and unknown variance σ^2

- Suppose we want to test null vs. alternative hypotheses

$$H_0 : \mu = \mu_0$$
$$H_a : \mu \neq \mu_0$$

  for some fixed value μ_0, which is what we are hypothesizing to be the unknown value of μ.

- Intuitively, we would expect that if $|\bar{X}_n - \mu_0|$ is large, then H_0 is not likely to be true.

- However, to develop test with known statistical properties, need a statistic (function of X_i's) whose distribution is known when H_0 is true.

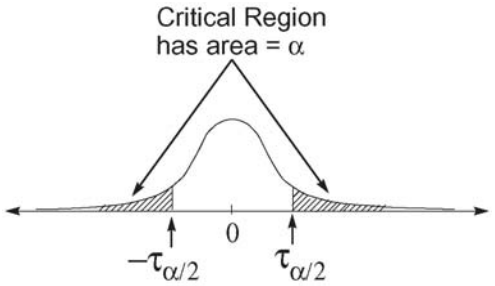
- If H_0 is true, then the test statistic

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2/n}} \sim t_{n-1}$$

- **Two-tailed hypothesis t-test** at level α for $H_0 : \mu = \mu_0$

If $|T| \geq \tau_{\alpha/2}$, then reject H_0,
If $|T| < \tau_{\alpha/2}$, then do not reject H_0,

where $\tau_{\alpha/2} \equiv \tau_{\alpha/2,n-1}$ satisfies $\mathcal{P}(t_{n-1} \geq \tau_{\alpha/2}) = \alpha/2$.



- **Remarks:**
  - Can show that reject H_0 at level α iff $\mu_0 \notin I$, where I is $100(1-\alpha)\%$ confidence interval for μ.
  - Can also define one-tailed hypothesis tests
- **Example:** Hypothesis test for mean return μ for INTC

$$H_0 : \mu = 0,$$
$$H_a : \mu \neq 0.$$

  - Under H_0, test statistic $T \sim t_{251}$, and for our data,

$$T = \frac{\bar{X}_n - \mu_0}{\sqrt{S_n^2/n}} = \frac{0.00315 - 0}{\sqrt{0.00054/252}} = 2.149$$

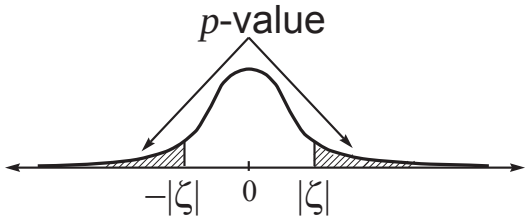
    and critical point $\tau_{0.025,\,251} = 1.969$.
  - Reject H_0 at level $\alpha = 0.05$ since $T \geq \tau_{0.025}$
  - 95% CI for μ is $I = (0.00026,\ 0.00604)$, and $0 \notin I$

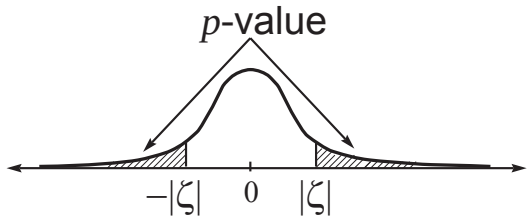
- When performing hypothesis test, two types of errors can arise

1. **Type I Error**: Rejecting H_0 when it is true.
   - Probability of Type I error = level α
   - Under experimenter's control.
2. **Type II Error**: Not rejecting H_0 when it is false
   - For fixed level α and sample size n, probability of Type II error depends on what is actually true (instead of H_0) and may be unknown.
   - β = probability of Type II error
   - **Power** of test, $\delta = 1 - \beta$, is probability of rejecting H_0 when it is false.

p-values

- Rather than specifying level α of test, can give p-value
- Suppose we have a test statistic T
- For the current data set, suppose value of T is ζ.
- **Definition:** p-value is probability (assuming H_0 is true) of observing at least as "unlikely" a value for test statistic T as ζ
  - For 2-tailed t-test, p-value $= \mathcal{P}(|T| \geq |\zeta|)$





- If p-value very small, then
  - test statistic highly unlikely if H_0 is true
  - so H_0 unlikely to be true
- If we want to perform hypothesis test at level α, then
  - reject H_0 if p-value $\leq \alpha$
  - do not reject H_0 if p-value $> \alpha$

Example: INTC mean return μ

- Hypotheses

$$H_0 : \mu = 0,$$
$$H_a : \mu \neq 0.$$

- When H_0 is true, test statistic

$$T = \frac{\sqrt{252}}{S_{252}} \bar{X}_{252} \sim t_{251}$$

- For our data $T = 2.149$
- What is probability of observing test statistic with value at least as extreme as 2.149 ?
  - p-value $= \mathcal{P}(|t_{251}| \geq |2.149|) = 0.016$
  - So reject H_0 at level $\alpha = 0.05$ since p-value ≤ 0.05.

Summary of Hypothesis Testing

- **Goal:** Make inferences about unknown distribution F
- **Approach:**
  - Form null hypothesis H_0 and alternative H_a.
  - Compute test statistic T from sample.
  - Specify critical region C for level α.
  - If $T \in C$, then reject H_0 at level α.
    Otherwise, do not reject H_0.
- Two types of errors can be made.
  - Type I error: rejecting H_0 when H_0 is true.
  - Type II error: not rejecting H_0 when H_0 is false.

- Connection between
  - $100(1-\alpha)\%$ confidence interval I for θ
  - level-α hypothesis test with $H_0 : \theta = \theta_0$
  - Reject H_0 iff $\theta_0 \notin I$
- Can also specify p-value
  - If performing level-α hypothesis test,
    reject H_0 if p-value $\leq \alpha$.
- Can also perform **one-sided** hypothesis test,
  - e.g., $H_0 : \mu \leq \mu_0$ vs. $H_a : \mu > \mu_0$ for some specified μ_0.
  - If data assumed to be normal and performing level-α t-test,
    - ▷ reject when test statistic $T \geq \tau_\alpha$;
    - ▷ do not reject when $T < \tau_\alpha$.
  - Note that Student t critical point is τ_α rather than $\tau_{\alpha/2}$.

- Can also perform hypothesis tests when data is not assumed to be normal.
  - Asymptotic validity typically justified by CLT.
  - To perform two-sided hypothesis test for mean μ, just replace Student t critical value $\tau_{\alpha/2}$ with normal critical value $z_{\alpha/2}$.