CENG 222 - Cheatsheet

Burak Metehan Tunçel - April 2022

1 Chapter 2: Probability

1.1 Events and Their Probabilities

Definitions

- **Outcome:** The result of certain random experiment.
- **Sample Space (Ω):** A collection of all (observable) outcomes, or elementary results, of an experiment. Not observable outcome in sample space.
- **Event:** Any set of outcomes.
- **Disjoint $\subseteq$ Mutually exclusive:** Events whose intersections is empty. $A_i \cap A_j = \emptyset$ for any $i \neq j$.
- **Exhaustive Events:** Events whose union is sample space. $A \cup B \cup C \cup \cdots = \Omega$.

Set Operations

- **Union ($\cup$):** `or`
- **Intersection ($\cap$):** `and`
- **Complement ($'$, A^c, $\overline{A}$):** `not`
- **Difference ($\setminus$):** `but not`
- $\overline{E_1 \cup \cdots \cup E_n} = \overline{E_1} \cap \cdots \cap \overline{E_n}$
  $\overline{E_1 \cap \cdots \cap E_n} = \overline{E_1} \cup \cdots \cup \overline{E_n}$

1.2 Rules of Probability

- $\boldsymbol{P}\{\Omega\} = 1$
- $\boldsymbol{P}\{\emptyset\} = 0$
- $\boldsymbol{P}\{E\} = \sum_{\omega_k \in E} \boldsymbol{P}\{\omega_k\}$ (Only for mutually exclusive events)
- $\boldsymbol{P}\{A \cup B\} = \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} - \boldsymbol{P}\{A \cap B\}$
- $\boldsymbol{P}\{A \cup B \cup C\} = \boldsymbol{P}\{A\} + \boldsymbol{P}\{B\} + \boldsymbol{P}\{C\} - \boldsymbol{P}\{A \cap B\} - \boldsymbol{P}\{A \cap C\} - \boldsymbol{P}\{B \cap C\} + \boldsymbol{P}\{A \cap B \cap C\}$
- $\boldsymbol{P}\{A\} + \boldsymbol{P}\{\overline{A}\} = \boldsymbol{P}\{A \cup \overline{A}\} = \boldsymbol{P}\{\Omega\} = 1$
- $\boldsymbol{P}\{E_1 \cap \cdots \cap E_n\} = \boldsymbol{P}\{E_1\} \cdot \ldots \cdot \boldsymbol{P}\{E_n\}$ (Independent Events)

1.3 Permutations and Combinations

- **Permutations with replacement:**

$$P_r(n,k) = \overbrace{n \cdot n \cdot \ldots \cdot n}^{\text{k terms}} = n^k$$

- **Permutations without replacement:**

$$P(n,k) = \overbrace{n(n-1)(n-2) \cdot \ldots \cdot (n-k+1)}^{\text{k terms}} = \frac{n!}{(n-k)!}$$

- **Combinations with replacement:**

$$C_r(n,k) = \binom{k+n-1}{k} = \frac{(k+n-1)!}{k!(n-1)!}$$

- **Combinations without replacement:**

$$C(n,k) = \binom{n}{k} = \frac{P(n,k)}{P(k,k)} = \frac{n!}{k!(n-k)!}$$

1.4 Conditional Probability and Independence

Definitions

- **Conditional Probability ($\boldsymbol{P}\{A \mid B\}$):** Conditional probability basically is “What is the probability of observing event A given that event B is observed?”
- **Independent Events:** Occurrence of one event does not affect the probabilities of other.

Formulas

- $\boldsymbol{P}\{A|B\} = \dfrac{\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{B\}}$ and $\boldsymbol{P}\{B|A\} = \dfrac{\boldsymbol{P}\{A \cap B\}}{\boldsymbol{P}\{A\}}$
- $\boldsymbol{P}\{A\}\boldsymbol{P}\{B \mid A\} = \boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{B\}\boldsymbol{P}\{A \mid B\}$
- $\boldsymbol{P}\{A \mid B\} = \boldsymbol{P}\{A\}$ and $\boldsymbol{P}\{B \mid A\} = \boldsymbol{P}\{B\}$ (Independent Events)
- $\boldsymbol{P}\{A \cap B\} = \boldsymbol{P}\{A\}\boldsymbol{P}\{B\}$ (Independent Events)
- $\boldsymbol{P}\{A|B\} = \dfrac{\boldsymbol{P}\{B|A\}\boldsymbol{P}\{A\}}{\boldsymbol{P}\{B\}}$ (Bayes Rule)
- $\boldsymbol{P}\{A\} = \sum_{j=1}^{k} \boldsymbol{P}\{A|B_j\}\boldsymbol{P}\{B_j\}$ (Law of Probability)
- $\boldsymbol{P}\{A|B\} = \dfrac{\boldsymbol{P}\{B|A\}\boldsymbol{P}\{A\}}{\boldsymbol{P}\{B|A\}\boldsymbol{P}\{A\} + \boldsymbol{P}\{B|\overline{A}\}\boldsymbol{P}\{\overline{A}\}}$ (Bayes Rule for Two Events)

2 Chapter 3: Discrete Random Variables and Their Distributions

2.1 Distribution of Random Variable

2.1.1 Main Concepts

- A **random variable** is a function of an outcome,

$$X = f(\omega)$$

  In other words, it is a quantity that depends on chance.

- Collection of all the probabilities related to X is the **distribution** of X.

- **Probability Mass Function**, or **PMF**:

$$P(x) = \boldsymbol{P}\{X = x\}$$

- **Cumulative Distribution Function**, or **CDF**:

$$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$$

- The set of possible values of X is called the **support** of the distribution F.

- For every outcome ω, the variable X takes one and only one value x. This makes events $\{X = x\}$ disjoint and exhaustive, and therefore,

$$\sum_{x} P(x) = \sum_{x} \boldsymbol{P}\{X = x\} = 1$$

2.1.2 Types of Random Variables

- *Discrete random variables*:

  These are variables whose range is *finite or countable.* In particular, it means that their values can be listed, or arranged in a sequence. Examples include the number of jobs submitted to a printer, the number of errors, the number of error-free modules, the number of failed components, and so on.

- *Continuous random variables*:

  This could be a bounded interval $(a,\ b)$, or an unbounded interval such as $(a,\ +\infty)$, $(-\infty,\ b)$, or $(-\infty,\ +\infty)$. Intervals are uncountable, therefore, all values of a random variable cannot be listed in this case. Examples of continuous variables include various times (software installation time etc., also physical variables like weight, height etc.

2.2 Distribution of a Random Vector

2.2.1 Joint Distribution and Marginal Distributions

If X and Y are random variables, then the pair $(X,\ Y)$ is a **random vector**. Its distribution is called the **joint distribution** of X and Y. Individual distributions of X and Y are then called the **marginal distributions**.

Similarly to a single variable, the *joint distribution* of a vector is a collection of probabilities for a vector $(X,\ Y)$ to take a value $(x,\ y)$. Recall that two vectors are equal,

$$(X,\ Y) = (x,\ y)$$

Addition Rule

$$P_X(x) = \boldsymbol{P}\{X = x\} = \sum_{y} P_{(X,\ Y)}(x,\ y)$$
$$P_Y(x) = \boldsymbol{P}\{Y = y\} = \sum_{x} P_{(X,\ Y)}(x,\ y)$$

2.2.2 Independence of Random Variables

Random variables X and Y are **independent** if

$$P_{X,\ Y}(x,\ y) = P_X(x)P_Y(y)$$

2.3 Expectation, Variance, Covariance and Correlation

2.3.1 Expectation

Expected value of a random variable X is its *mean, the average value.*

$$\mu = \mathbf{E}(X) = \sum_{x} xP(x)$$
$$\mu = \mathbf{E}(g(x)) = \sum_{x} g(x)P(x)$$

2.3.2 Variance and Standard Deviation

Variance, the *expected squared deviation* from the mean.

$$\begin{aligned}\sigma^2 = \mathrm{Var}(X) &= \mathbf{E}(X - \mathbf{E}(X))^2 \\ &= \mathbf{E}(X - \mu)^2 \\ &= \sum_{x} (x - \mu)^2 P(x)\end{aligned}$$

Standard deviation, the *Square root of variance*,

$$\sigma = \mathrm{Std}(X) = \sqrt{\mathrm{Var}(X)}$$

2.3.3 Covariance and Correlation

Covariance: $\sigma_{XY} = \mathrm{Cov}(X,\ Y)$ is defined as

$$\begin{aligned}\mathrm{Cov}(X,\ Y) &= \mathbf{E}\{(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))\} \\ &= \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)\end{aligned}$$

It summarizes interrelation of two random variables.

Correlation coefficient: Between variables X and Y is defined as

$$\rho = \frac{\mathrm{Cov}(X,\ Y)}{(\mathrm{Std}X)(\mathrm{Std}Y)}$$

Correlation coefficient is a rescaled, normalized covariance.

2.3.4 Properties

- **Expectation:**

$$\mathbf{E}(aX + bY + c) = a\mathbf{E}(X) + b\mathbf{E}(Y) + c$$

For *independent* X and Y,

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$$

- **Variance:**

$$\begin{aligned}\mathrm{Var}(aX + bY + c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \\ + 2ab\mathrm{Cov}(X,\ Y)\end{aligned}$$

In particular,

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

For *independent* X and Y,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

- **Covarince**:
  - If $\mathrm{Cov}(X,\ Y) > 0$, these variables are ***positively correlated.***
  - If $\mathrm{Cov}(X,\ Y) < 0$, these variables are ***negatively correlated.***
  - If $\mathrm{Cov}(X,\ Y) = 0$, these variables are ***uncorrelated.***

$$\begin{aligned}\mathrm{Cov}(aX + bY,\ cZ + dW) &= ac\mathrm{Cov}(X,\ Z) \\ &+ ad\mathrm{Cov}(X,\ W) \\ &+ bc\mathrm{Cov}(Y,\ Z) \\ &+ bd\mathrm{Cov}(Y,\ W) \\ \mathrm{Cov}(X,\ Y) &= \mathrm{Cov}(Y,\ X)\end{aligned}$$

In particular,

$$\mathrm{Cov}(aX + b,\ cY + d) = ac\mathrm{Cov}(X,\ Y)$$

For *independent* X and Y,

$$\mathrm{Cov}(X,\ Y) = 0$$

- **Correlation**:
  -

$$-1 \le \rho \le 1$$

where $|\rho| = 1$ is possible only when all values of X and Y lie on a straight line.

  - If $\rho = 1$, it is ***strong (perfect) positive correlation.***
  - If $\rho = -1$, it is ***strong (perfect) negative correlation.***
  - If $\rho = 0$, it is ***weak or no correlation.***

$$\rho(X,\ Y) = \rho(Y,\ X)$$

In particular,

$$\rho(aX + b,\ cY + d) = \rho(X,\ Y)$$

2.3.5 General Notation

$$\begin{aligned}\mu \text{ or } \mathbf{E}(X) &= \text{expectation} \\ \sigma_X^2 \text{ or } \mathrm{Var}(X) &= \text{variance} \\ \sigma_X \text{ or } \mathrm{Std}(X) &= \text{standard deviation} \\ \sigma_{XY} \text{ or } \mathrm{Cov}(X,\ Y) &= \text{covariance} \\ \rho_{XY} &= \text{correlation coefficient}\end{aligned}$$

The following page includes the "Families of Discrete Distribution". The following box is the a little summary.

Summary of Families of Discrete Distribution

Distribution	Number of	In/For
Binomial	Successes	In n trials
Geometric	Trials	For first success
Negative Binomial	Trials	For k successes
Poisson	Rare events	In fixed time

Additionally, in Binomial distribution, if $n \ge 30$ and $p \le 0.05$, then

$$\texttt{Binomial}(n, p) \approx \texttt{Poisson}(\lambda) \quad (\lambda = np)$$

2.4 Families of Discrete Distributions

2.4.1 Bernoulli Distribution

Definitions

- **Bernoulli variable**: A random variable with two possible values, 0 and 1.
- **Bernoulli distribution**: Distribution of *bernoulli variable.*
- **Bernoulli trial**: Any experiment with a *binary outcome.*

Summary of Bernoulli Distribution

$$\begin{aligned} p &= \text{probability of success} \\ q &= \text{probability of failure} \\ P(x) &= \begin{cases} q = 1-p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases} \\ \mathbf{E}(X) &= p \\ \mathrm{Var}(X) &= pq \quad [= p(1-p)] \end{aligned}$$

2.4.2 Binomial Distribution

Number of successes *in a sequence of independent Bernoulli* ***trials*** has Binomial distribution.

Summary Binomial Distribution

$$\begin{aligned} n &= \text{number of trials} \\ p &= \text{probability of success} \\ P(x) &= \binom{n}{x} p^x q^{n-x} \\ \mathbf{E}(X) &= np \\ \mathrm{Var}(X) &= npq = np(1-p) \end{aligned}$$

2.4.3 Geometric Distribution

The number of Bernoulli trials *needed to get the* ***first success*** has Geometric distribution.

Summary of Geometric Distribution

$$\begin{aligned} p &= \text{probability of success} \\ P(x) &= (1-p)^{x-1}p, \quad x = 1, 2, \dots \\ \mathbf{E}(X) &= \frac{1}{p} \\ \mathrm{Var}(X) &= \frac{1-p}{p^2} \end{aligned}$$

2.4.4 Negative Binomial Distribution

Sequence of independent Bernoulli trials, the ***number of trials*** *needed to* ***obtain*** k ***successes*** has Negative Binomial distribution.

Summary of Negative Binomial Distribution

$$\begin{aligned} k &= \text{number of success} \\ p &= \text{probability of success} \\ P(x) &= (1-p)^{x-k}p^k, \quad x = k, k+1, \dots \\ \mathbf{E}(X) &= \frac{k}{p} \\ \mathrm{Var}(X) &= \frac{k(1-p)}{p^2} \end{aligned}$$

2.4.5 Poisson Distribution

The ***number of rare events*** *occurring within a* ***fixed period of time*** has Poisson distribution.

Summary of Poisson Distribution

$$\begin{aligned} \lambda &= \text{frequency,} \\ &\quad \text{average number of events} \\ P(x) &= e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 1, 2, \dots \\ \mathbf{E}(X) &= \lambda \\ \mathrm{Var}(X) &= \lambda \end{aligned}$$

2.4.6 Poisson Approximation of Binomial Distribution

If

$$n \geq 30 \qquad\qquad p \leq 0.05$$

Poisson distribution can be effectively used to approximate Binomial probabilities. (If $p \geq 0.95$, then use $q \leq 0.05$ and consider failure instead of success.)

Summary of Poisson Approximation of Binomial Distribution

$$\mathtt{Binomial}(n, p) \approx \mathtt{Poisson}(\lambda)$$

where $n \geq 30$, $p \leq 0.05$, $np = \lambda$.

Remark: Mathematically, it means closeness of Binomial and Poisson pmf

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}$$

3 Chapter 4: Continuous Distributions

3.1 Probability Density

3.1.1 Definitions

- Probability mass function (PMF), $P(x)$, is always equal to 0.
- Cumulative Distribution Function (CDF), $F(x)$, is useful.
- Probability Density Function (PDF, density), $f(x)$ is the derivative of the cdf. The distribution is called continuous if it has a density.

3.1.2 Notation

- **PMF:** $P(x)$
- **CDF:** $F(x)$
- **PDF:** $f(x)$

3.1.3 Formulas and Equations

- $P(x) = 0$ for all x.
- $F(x) = \boldsymbol{P}\{X \leq x\} = \boldsymbol{P}\{X < x\}$
- $P(x) = \boldsymbol{P}\{x \leq X \leq x\} = \int_x^x f = 0$
- $f(x) = F'(x)$.
- $\int_a^b f(x)dx = F(b) - F(a) = \boldsymbol{P}\{a < X < b\}$
  - $\int_{-\infty}^b f(x)dx = \boldsymbol{P}\{-\infty < X < b\} = F(b)$
  - $\int_{-\infty}^\infty f(x)dx = \boldsymbol{P}\{-\infty < X < \infty\} = 1$

3.1.4 Analogy: PMF vs. PDF

Distribution	**Discrete**	**Continuous**
Definition	$P(x) = \boldsymbol{P}\{X = x\}$	$f(x) = F'(x)$
Computing probabilities	$\boldsymbol{P}\{X \in A\} = \sum_{x \in A} P(x)$	$\boldsymbol{P}\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = \boldsymbol{P}\{X \leq x\} = \int_{-\infty}^x f(y)dy$
Total probability	$\sum_x P(x) = 1$	$\int_{-\infty}^{+\infty} f(x)dx = 1$
Marginal Distributions	$P(x) = \sum_y P(x, Y)$ $P(y) = \sum_x P(x, Y)$	$f(x) = \int f(x,y)dy$ $f(y) = \int f(x,y)dx$
Independence	$P(x,y) = P(x)P(y)$	$f(x,y) = f(x)f(y)$
Computing Probabilities	$\boldsymbol{P}\{(X,Y) \in A\} = \sum\sum_{(x,y) \in A} P(x,y)$	$\boldsymbol{P}\{(X,Y) \in A\} = \int\int_{(x,y) \in A} f(x,y)dxdy$

Table 1: **Row 1-4:** *PMF $P(x)$ vs. PDF $f(x)$* | **Row 5-7:** *Joint and Marginal Distributions*

3.1.5 Expectation and Variance

Discrete	**Continuous**
$\mathbf{E}(X) = \sum_x xP(x)$	$\mathbf{E}(X) = \int xf(x)dx$
$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2 = \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2 = \int (x-\mu)^2 f(x)dx$ $= \int x^2 f(x)dx - \mu^2$
$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x\sum_y (x-\mu_X)(y-\mu_Y)P(x,y)$ $= \sum_x\sum_y (xy)P(x,y) - \mu_X\mu_Y$	$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \int\int (x-\mu_X)(y-\mu_Y)f(x,y)dxdy$ $= \int\int (xy)f(x,y)dxdy - \mu_X\mu_Y$

Table 2: *Moments for discrete and continuous distributions.*

3.2 Families of Continuous Distribution

3.2.1 Uniform Distribution

Outcomes with equal density. Picking "random" from a given interval; that is, *without any preference* to lower, higher, or medium values.

$$f(x) = \frac{1}{b-a}, \qquad a \leq x \leq b$$

Uniform Property The probability is only determined by the length of the interval, but not by its location.

- $\boldsymbol{P}\{t < X < t+h\} = \int_t^{t+h} \frac{1}{b-a} dx = \frac{h}{b-a}$
- $\boldsymbol{P}\{s < X < s+h\} = \boldsymbol{P}\{t < X < t+h\}$

Standard Uniform Distribution The Uniform distribution with $a = 0$ and $b = 1$ is called *Standard Uniform distribution.* The Standard Uniform density is $f(x) = 1$ for $0 < x < 1$.

Let X be a Uniform(a, b) random variable (*not standard uniform random variable*), then

$$Y = \frac{X-a}{b-a}$$

is Standard Uniform. Likewise, Let Y be Standard Uniform, then

$$X = a + (b-a)Y$$

is Uniform(a, b).

Summary of Uniform Distribution

$$\begin{aligned} (a, b) &= \text{range of values} \\ f(x) &= \frac{1}{b-a}, \qquad a \leq x \leq b \\ \mathbf{E}(X) &= \frac{b+a}{2} \\ \mathrm{Var}(X) &= \frac{(b-a)^2}{12} \end{aligned}$$

3.2.2 Exponential Distribution

Time between events, time until an event. In a sequence of rare events, *when the number of events is Poisson, the time between events is Exponential.*

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x} && \text{for } x > 0 \\ \boldsymbol{P}\{X \leq x\} = F(x) &= 1 - e^{-\lambda x} && (x > 0) \end{aligned}$$

Important Note

$$\boldsymbol{P}\{X > x\} = 1 - \boldsymbol{P}\{X \leq x\} = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

In here X is "*time between two events*" and x is "*particular time*".

"***Exponential distribution*** is a continuous version of ***Geometric distribution***. In the *Geometric distribution*, we analyze the how many trials are required before the first success, whereas, in the *Exponential distribution*, we analyze the time until the event."

Memoryless Property Memoryless property means that the fact of having waited for t minutes gets "*forgotten*", and it *does not affect the future waiting time.* Mathematically,

$$\boldsymbol{P}\{T > t + x \mid T > t\} = \boldsymbol{P}\{T > x\} \qquad \text{for } t, x > 0$$

In this formula, t is the already elapsed portion of waiting time, and x is the additional, remaining time.

This property is unique for Exponential distribution. No other continuous variable $X \in (0, \infty)$ is memoryless. Among discrete variables, such a property belongs to Geometric distribution.

Summary of Exponential Distribution

$$\begin{aligned} \lambda &= \text{frequency parameter,} \\ &\quad \text{the number of events per time unit} \\ f(x) &= \lambda e^{-\lambda x}, \qquad x \geq 0 \\ \mathbf{E}(X) &= \frac{1}{\lambda} \\ \mathrm{Var}(X) &= \frac{1}{\lambda^2} \end{aligned}$$

3.2.3 Gamma Distribution

α ***independent steps***, and each step takes ***Exponential(*λ*)*** amount of time, the total time has Gamma distribution with parameters α and λ.

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

$$F(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^t x^{\alpha-1} e^{-\lambda x} dx$$

where $\Gamma(\alpha) = (\alpha - 1)!$

Special Cases

$$\text{Gamma}(1, \lambda) = \text{Exponential}(\lambda)$$
$$\text{Gamma}(\alpha, 1/2) = \text{Chi-square}(2\alpha)$$

Summary of Gamma Distribution

$$\begin{aligned}
\alpha &= \text{shape parameter} \\
\lambda &= \text{frequency parameter} \\
f(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0 \\
\mathbf{E}(X) &= \frac{\alpha}{\lambda} \\
\text{Var}(X) &= \frac{\alpha}{\lambda^2}
\end{aligned}$$

Gamma-Poisson Formula Let T be a Gamma variable with an integer parameter α and some positive λ. This is a distribution of the time of the α-th rare event. Then, the event $\{T > t\}$ means that the α-th rare event occurs after the moment t, and therefore, *fewer than α rare events occur before the time t*. We see that

$$\{T > t\} = \{X < \alpha\}$$

where X is the number of events that occur before the time t. This number of rare events X has Poisson distribution with parameter (λt) [*poisscdf($X, \lambda t$)*]; therefore, the probability

$$\boldsymbol{P}\{T > t\} = \boldsymbol{P}\{X < \alpha\}$$

and the probability of a complement

$$\boldsymbol{P}\{T \le t\} = \boldsymbol{P}\{X \ge \alpha\}$$

can both be computed using the Poisson distribution of X.

Summary of Gamma-Poisson Formula

For a Gamma(α, λ) variable T and a Poisson(λt) variable X,

$$\boldsymbol{P}\{T > t\} = \boldsymbol{P}\{X < \alpha\}$$
$$\boldsymbol{P}\{T \le t\} = \boldsymbol{P}\{X \ge \alpha\}$$

3.2.4 Normal Distribution

Values with a bell-shaped distribution.

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

where parameters μ and σ have a simple meaning of the expectation $\mathbf{E}(X)$ and the standard deviation Std(X).

This density is known as the bell-shaped curve, symmetric and centered at μ, its spread being controlled by σ. Changing μ shifts the curve to the left or to the right without affecting its shape, while changing σ makes it more concentrated or more flat.

Summary of Normal Distribution

$$\begin{aligned}
\mu &= \text{expectation,} \\
&\quad \textit{location} \text{ parameter} \\
\sigma &= \text{standard deviation,} \\
&\quad \textit{scale} \text{ parameter} \\
f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < +\infty \\
\mathbf{E}(X) &= \mu \\
\text{Var}(X) &= \sigma^2
\end{aligned}$$

Standard Normal Distribution Normal distribution with "standard parameters" $\mu = 0$ and $\sigma = 1$.

Notation:

$$\begin{aligned}
Z &= \text{Standard Normal random variable} \\
\phi(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{ Standard Normal pdf} \\
\Phi(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \text{ Standard Normal cdf}
\end{aligned}$$

A Standard Normal variable, usually denoted by Z, can be obtained from a non-standard Normal(μ, σ) random variable X by *standardizing*, that is, subtracting the mean and dividing by the standard deviation,

$$Z = \frac{X - \mu}{\sigma}$$

Unstandardizing Z, we can reconstruct the initial variable X,

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

3.3 Central Limit Theorem

Theorem 1: Central Limit Theorem

Let $X_1, X_2, ...$ be independent random variables with the same expectation $\mu = \mathbf{E}(X_i)$ and the same standard deviation $\sigma = \text{Std}(X_i)$, and let

$$S_n = \sum_{i=1}^{n} X_i = X_i + ... + X_n$$

As $n \to \infty$, the standardized sum

$$Z_n = \frac{S_n - \mathbf{E}(S_n)}{\text{Std}(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a Standard Normal random variable, that is,

$$F_{Z_n}(z) = \boldsymbol{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right\} \to \Phi(z)$$

for all z.

As long as n is large (the rule of thumb is $n > 30$), one can use **Normal distribution** to compute probabilities about S_n.

Among the random variables, at least three have a form of S_n:

Binomial variable	=	sum of independent Bernoulli variables
Negative Binomial variable	=	sum of independent Geometric variables
Gamma variable	=	sum of independent Exponential variables

Hence, the Central Limit Theorem applies to all these distributions with sufficiently large n in the case of Binomial, k for Negative Binomial, and α for Gamma variables.

3.3.1 Normal Approximation to Binomial Distribution

Binomial variables represent a special case of

$$S_n = X_1 + ... + X_n$$

where all X_i have Bernoulli distribution with some parameter p.

- Small or large $p \Rightarrow$ approximate Binomial distribution with Poisson
- Moderate values of p (say, $0.05 \leq p \leq 0.95$) and for large n, `Central Limit Theorem` can be used

$$\text{Binomial}(n, p) \approx Normal\left(\mu = np, \sigma = \sqrt{np(1-p)}\right)$$

3.3.2 Continuity Correction

- This correction is needed when we approximate a discrete distribution (Binomial in this case) by a continuous distribution (Normal). Expand the interval by 0.5 units in each direction, then use the Normal approximation. Notice that

  $$P_X(x) = \boldsymbol{P}\{X = x\} = \boldsymbol{P}\{x - 0.5 < X < x + 0.5\}$$

  is true for a Binomial variable X; therefore, the continuity correction does not change the event and preserves its probability. Every time when we approximate some discrete distribution with some continuous distribution, we should be using a continuity correction.
- When a continuous distribution (say, Gamma) is approximated by another continuous distribution (Normal), the continuity correction is not needed. In fact, it would be an error to use it in this case because it would no longer preserve the probability.

4 Chapter8: Introdcution to Statistics

NOTATION

$$\begin{aligned}
\mu &= \text{population mean} \\
\bar{X} &= \text{Sample mean, estimator of } \mu \\
\\
\sigma &= \text{population standard deviation} \\
s &= \text{sample standard deviation, estimator of } \sigma \\
\\
\sigma^2 &= \text{population variance} \\
s^2 &= \text{sample variance, estimator of } \sigma \\
\\
\sigma(\hat{\theta}) &= \text{standard error of estimator } \hat{\theta} \text{ of parameter } \theta \\
s(\hat{\theta}) &= \text{estimated standard error} = \hat{\sigma}(\hat{\theta})
\end{aligned}$$

4.1 Mean

Population Mean: $\mu = \mathbf{E}(X)$

Sample Mean

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

4.1.1 Unbiasedness

Unbiased

$$\mathbf{E}\left(\hat{\theta}\right) = \theta$$

Bias of $\hat{\theta}$ is defined as $\text{Bias}(\hat{\theta}) = \mathbf{E}\left(\hat{\theta} - \theta\right)$.

4.2 Variance and Standard Deviation

Sample Variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Another formula

$$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

Sample Standard Deviation

$$s = \sqrt{s^2}$$

Variance of $\bar{X}$

$$\text{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

Standard Deviation of $\bar{X}$

$$\sigma(\bar{X}) = \sqrt{\text{Var}\left(\bar{X}\right)} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

4.3 Standard Errors of Estimates

Standard error of an estimator $\hat{\theta}$ is its standard deviation, $\sigma(\hat{\theta}) = \mathrm{Std}(\hat{\theta})$