CS145: Lecture 9 Outline

- Continuous Random Variables & Probability Densities
- Gaussian (Normal) Distributions
Cumulative Distribution Function (CDF)

- Recall probability mass function (PMF):
  \[ p_X(x) = P(X = x) \]

- The cumulative distribution function (CDF) is the cumulative sum of the PMF:
  \[ F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k) \]

- The CDF equals 0 below the range of X, 1 above the range of X, and is monotonically increasing:
  \[ F_X(x_2) \geq F_X(x_1) \text{ if } x_2 > x_1 \]

- The CDF allows quick computation of the probability of intervals:
  \[ P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \]
Examples of Discrete PMFs & CDFs

Uniform

$$n = b - a + 1$$

Geometric

$$P(X = x)$$

Binomial

$$P(X \leq x)$$
Continuous Random Variables

- For any discrete random variable, the CDF is \textit{discontinuous and piecewise constant}.
- If the CDF is \textit{monotonically increasing and continuous}, have a continuous random variable:
  \[ 0 \leq F_X(x) \leq 1 \]
  \[ F_X(x_2) \geq F_X(x_1) \text{ if } x_2 > x_1. \]
  \[ \lim_{x \to -\infty} F_X(x) = 0 \quad \lim_{x \to +\infty} F_X(x) = 1 \]
- The probability that continuous random variable \( X \) lies in the interval \((x_1, x_2]\) is then
  \[ P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \]
Continuous CDFs Define Probability Laws

Continuous random variables satisfy the **axioms of probability**.

**Non-negativity:**

\[ P(x_1 < X \leq x_2) \geq 0 \] for any \( x_1, x_2 \).

**Normalization:**

\[ P(-\infty < X < +\infty) = F_X(+\infty) - F_X(-\infty) = 1 - 0 = 1. \]

**Additivity:**

If \( x_1 < x_2 < x_3 \),

\[ P(x_1 < X \leq x_3) = P(x_1 < X \leq x_2) + P(x_2 < X \leq x_3) \]

\[ F_X(x_3) - F_X(x_1) = (F_X(x_2) - F_X(x_1)) + (F_X(x_3) - F_X(x_2)) \]

- The probability that continuous random variable \( X \) lies in interval \((x_1, x_2]\) is

\[ P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \]
If the CDF is differentiable, its first derivative is called the *probability density function (PDF)*:

\[ f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x) \]

By the *fundamental theorem of calculus*:

\[ \int_{x_1}^{x_2} f_X(x) \, dx = F_X(x_2) - F_X(x_1) = P(x_1 < X \leq x_2) \]

For any valid PDF:

\[ f_X(x) \geq 0 \]

\[ \int_{-\infty}^{+\infty} f_X(x) \, dx = 1 \]

\[ 0 \leq \int_{x_1}^{x_2} f_X(x) \, dx \leq 1 \]
Continuous Uniform Probability Densities

- Relationship between CDF and PDF:
  \[ F_X(x) = \int_{-\infty}^{x} f_X(t) \, dt \quad f_X(x) = \frac{dF_X(x)}{dx} \]

- For a continuous uniform random variable, an interval’s probability is proportional to length:
  
  If \( a \leq x_1 < x_2 \leq b \),
  
  \[ P(x_1 < X \leq x_2) = \frac{x_2 - x_1}{b - a} \]

- Note that it is possible that \( f_X(x) > 1 \)
  
  If \( a = 0 \) and \( b = 0.1 \), then \( \frac{1}{b - a} = 10 \).
General Continuous Random Variables

![Diagram of a continuous random variable with PDF](image)

\[ f_X(x) \geq 0 \]

\[
P(a \leq X \leq b) = \int_a^b f_X(x) \, dx \]  

\[
P(x \leq X \leq x + \delta) = \int_x^{x+\delta} f_X(s) \, ds \approx f_X(x) \cdot \delta
\]

**Observation:** For a continuous random variable, the probability of observing \(X = x\) for any particular real number \(x\) equals zero:

\[
P(X = x) = \lim_{\delta \to 0} P(x - \delta < X \leq x) = \lim_{\delta \to 0} \int_{x-\delta}^{x} f_X(s) \, ds = 0
\]

As floating point precision increases, probability of any particular number decreases.
General Continuous Random Variables

\[ f_X(x) \geq 0 \]

\[ P(a \leq X \leq b) = \int_a^b f_X(x) \, dx \]

\[ P(x \leq X \leq x + \delta) = \int_x^{x+\delta} f_X(s) \, ds \approx f_X(x) \cdot \delta \]

**Observation:** A PDF may take on arbitrarily large positive values:

\[ f_X(x) = \begin{cases} 
\frac{1}{2\sqrt{x}} & \text{if } 0 < x \leq 1, \\
0 & \text{otherwise.}
\end{cases} \]

\[ \int_{-\infty}^{\infty} f_X(x) \, dx = \int_0^1 \frac{1}{2\sqrt{x}} \, dx = \sqrt{x} \bigg|_0^1 = 1. \]
Expectations of Continuous Variables

- The *expectation* or *expected value* of a continuous random variable is:

\[ E[X] = \int_{-\infty}^{+\infty} x f_X(x) \, dx \]

- The expected value of a function of a continuous random variable:

\[ E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) \, dx \]

- The variance of a continuous random variable:

\[ \text{Var}[X] = E[X^2] - E[X]^2 = E[(X - E[X])^2] = \int_{-\infty}^{+\infty} (x - E[X])^2 f_X(x) \, dx \]

- **Intuition**: Create a discrete variable by quantizing \( X \), and compute discrete expectation. As number of discrete values grows, sum approaches integral.
The expectation of $X$ is

$$E[X] = \int_{a}^{b} \frac{x}{b-a} \, dx = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2},$$

and the second moment is

$$E[X^2] = \int_{a}^{b} \frac{x^2}{b-a} \, dx = \frac{b^3 - a^3}{3(b-a)} = \frac{b^2 + ab + a^2}{3}.$$

The variance is computed by

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{b^2 + ab + a^2}{3} - \frac{(b+a)^2}{4}$$

$$= \frac{(b-a)^2}{12}.$$
Conditioning a Uniform Distribution

Lemma

Let $X$ be a uniform random variable on $[a, b]$. Then for $c \leq d$

$$\Pr(X \leq c \mid X \leq d) = \frac{c - a}{d - a}.$$  

That is, conditioned on the fact that $X \leq d$, $X$ is uniform on $[a, d]$.

Proof.

$$\Pr(X \leq c \mid X \leq d) = \frac{\Pr((X \leq c) \cap (X \leq d))}{\Pr(X \leq d)}$$

$$= \frac{\Pr(X \leq c)}{\Pr(X \leq d)}$$

$$= \frac{c - a}{d - a}.$$
Exponential Distribution

Definition

The exponential distribution with parameter $\theta$:

$$F(x) = \begin{cases} 1 - e^{-\theta x} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$  

$$f(x) = \theta e^{-\theta x}, \text{ for } x \geq 0.$$  

$$E[X] = \int_0^\infty t\theta e^{-\theta t} dt = \frac{1}{\theta}.$$  

$$E[X^2] = \int_0^\infty t^2\theta e^{-\theta t} dt = \frac{2}{\theta^2}.$$  

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{1}{\theta^2}.$$
With the matched exponential and geometric parameters given above:

\[ F_{\text{exp}}(n\delta) = F_{\text{geo}}(n), \quad n = 1, 2, \ldots, \]

Interpretation: If we very quickly toss a coin (every \( \delta \ll 1 \) seconds) toss a coin with a very small probability of coming up heads, the distribution of the time until the first head is approximately exponential.
Exponential Distributions are Memoryless

Lemma

For an exponential random variable with parameter $\theta$, 

$$
\Pr(X > s + t \mid X > t) = \Pr(X > s)
$$

Proof.

$$
\Pr(X > s + t \mid X > t) = \frac{\Pr(X > s + t)}{\Pr(X > t)}
= \frac{1 - \Pr(X \leq s + t)}{1 - \Pr(X \leq t)}
= \frac{e^{-\theta(s+t)}}{e^{-\theta t}}
= e^{-\theta s} = \Pr(X > s).
$$
CS145: Lecture 9 Outline

- Continuous Random Variables & Probability Densities
- Gaussian (Normal) Distributions
Gaussian (Normal) Distributions

The density function of the **Normal distribution** $N(\mu, \sigma^2)$ is:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}$$

The distribution function:

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2} dt$$

Properties:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} dx = 1.$$  

$$E(X) = \mu \quad \text{Var}(X) = \sigma^2$$

The integral has no closed form.
Why the Normal Distribution?

**Empirical observation:** Many random phenomena follow (at least approximately) Normal distribution.

- Height, weight, income,....
- The velocity of molecule in gas (Brownian Motion)
- Measurement error, noise...
- ....

**The Central Limit Theorem:**
“The distribution of the average of large number of independent random variable converges to the Normal distribution”.

**Binomial Distribution:**

\[ p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n \]
Scaling a Gaussian Variable

\[ f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} \]

\[ E[X] = \mu \]

\[ \text{Var}[X] = E[(X - \mu)^2] = \sigma^2 \]

\[ \sqrt{\text{Var}[X]} = \sigma \] is the standard deviation

Any linear transformation of a Gaussian variable is Gaussian!

\[ Y = aX + b \quad f_Y(y) = \frac{1}{\sqrt{2\pi\bar{\sigma}^2}} e^{-\frac{1}{2} \left( \frac{y - \bar{\mu}}{\bar{\sigma}} \right)^2} \]

- Mean and variance of linear functions: \( \bar{\mu} = a\mu + b, \quad \bar{\sigma} = |a|\sigma \)
- Proof that PDF is Gaussian will come later in course…
Standard Normal Random Variables

- If \( X \sim N(\mu, \sigma^2) \) then for any constants \( a \) and \( b \) the random variable \( aX + b \) is distributed \( N(a\mu + b, a^2\sigma^2) \).
- If \( X \sim N(\mu, \sigma^2) \) then \( Z = \frac{X - \mu}{\sigma} \) is distribution \( N(0, 1) \).
- \( N(0, 1) \) is the standard Normal distribution.

\[
\Pr(Z \leq z) = \Phi_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} \, dt
\]

\[
\phi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}
\]
Classic Computation of Normal CDF

• If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X - \mu}{\sigma}$

• $Pr(X \leq x) = Pr\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$

• The standard Normal random variable is symmetric around 0.

• For $\frac{x - \mu}{\sigma} < 0$,

$$\Phi\left(\frac{x - \mu}{\sigma}\right) = 1 - \Phi\left(-\frac{x - \mu}{\sigma}\right)$$

• With a table of $\Phi(Z)$ for $Z > 0$ we can compute $F(x)$ for any Normal random variable
Modern Computation of Normal CDF

**normcdf**

Normal cumulative distribution function

**Syntax**

\[
p = \text{normcdf}(x)
\]

\[
p = \text{normcdf}(x, \mu, \sigma)
\]

\[
[p, pl, pu] = \text{normcdf}(x, \mu, \sigma, pco, \alpha)
\]

\[
[p, pl, pu] = \text{normcdf}([], 'upper')
\]

**Description**

\(p = \text{normcdf}(x)\) returns the standard normal cdf at each value in \(x\). The standard normal distribution has parameters \(\mu = 0\) and \(\sigma = 1\). \(x\) can be a vector, matrix, or multidimensional array.

\(p = \text{normcdf}(x, \mu, \sigma)\) returns the normal cdf at each value in \(x\) using the specified values for the mean \(\mu\) and standard deviation \(\sigma\). \(x, \mu, \) and \(\sigma\) can be vectors, matrices, or multidimensional arrays that all have the same size. A scalar input is expanded to a constant array with the same dimensions as the other inputs. The parameters in \(\sigma\) must be positive.

**norminv**

Normal inverse cumulative distribution function

**Syntax**

\[
X = \text{norminv}(P, \mu, \sigma)
\]

\[
[X, XL, XU] = \text{norminv}(P, \mu, \sigma, pco, \alpha)
\]

**Description**

\(X = \text{norminv}(P, \mu, \sigma)\) computes the inverse of the normal cdf using the corresponding mean \(\mu\) and standard deviation \(\sigma\) at the corresponding probabilities in \(P\). \(P, \mu, \) and \(\sigma\) can be vectors, matrices, or multidimensional arrays that all have the same size. A scalar input is expanded to a constant array with the same dimensions as the other inputs. The parameters in \(\sigma\) must be positive, and the values in \(P\) must lie in the interval \([0, 1]\).