Random Variables

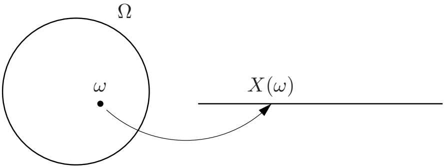
Lecture 2

Aug. 28, and Sep. 2 and 4, 2025

Outline

Random Variable

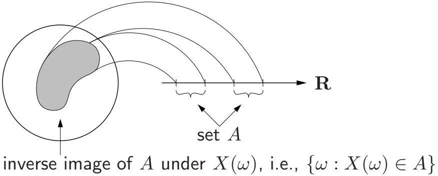
- A **random variable** (r.v.) is a real-valued function $X(\omega)$ over a sample space Ω, i.e., $X : \Omega \to \mathbb{R} = (-\infty, \infty)$. Formally, given a probability space $(\Omega, \mathcal{F}, P)$, $X : \Omega \to \mathbb{R}$ is such that $\{\omega \,|\, X(\omega) \leq x\} = \{X \leq x\}$ is an event (belongs to $\mathcal{F}$) for any real value x, and $P(X = -\infty) = P(X = \infty) = 0$.



- Notation
  - We use upper case letters for random variables X, Y, Θ, $\cdots$.
  - We use lower case letters x, y, θ, $\cdots$ for **values** of random variables: $X = x$ means that random variable X takes on the value x, i.e., $X(\omega) = x$ where ω is the outcome in Ω.

Specifying a Random Variable

- Specifying a random variable means being able to determine the probability that $X \in A$ for any Borel set $A \subset \mathbb{R}$, in particular, for any interval $(a, b]$.
- To do so, consider the inverse image (back to Ω) of A under X, i.e., $\{\omega \,|\, X(\omega) \in A\}$



- Since $X \in A$ if and only if (iff) $\omega \in \{\omega \,|\, X(\omega) \in A\}$,

$$P(\{X \in A\}) = P(\{\omega \,|\, X(\omega) \in A\}) = P(X \in A)$$

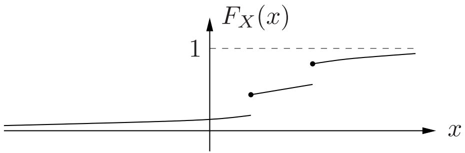
Shorthand notation: $P(\{\text{set description}\}) = P(\text{set description})$

Cumulative Distribution Function (CDF)

- We need to be able to determine $P(X \in A)$ for any Borel set $A \subset \mathbb{R}$, i.e., any set generated by starting from intervals and taking countable unions, intersections, and complements.
- Hence, it suffices to specify $P(X \in (a, b])$ for all intervals. The probability of any other Borel set can be determined by the axioms of probability.
- Equivalently, it suffices to specify its **cumulative distribution function** (cdf):

$$F_X(x) \stackrel{\text{def}}{=} P(X \leq x) = P(X \in (-\infty, x]), \quad x \in \mathbb{R}$$

- Properties of cdf:
  - $F_X(x) \geq 0$, $-\infty < x < \infty$
  - $F_X(x)$ is nondecreasing, i.e., if $a > b$ then $F_X(a) \geq F_X(b)$



- For $a > b$, since $\{X \le a\} = \{X \le b\} \cup \{b < X \le a\}$, we have

$$F_X(a) = F_X(b) + \underbrace{P(\{b < X \le a\})}_{\ge 0} \ge F_X(b)$$

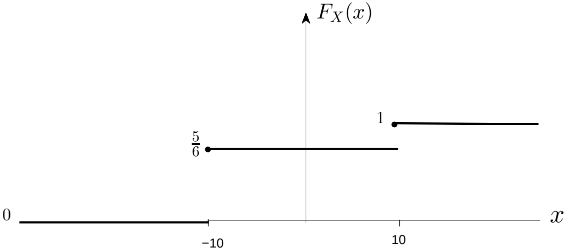
- Limits: $\lim_{x\to\infty} F_X(x) = 1$ and $\lim_{x\to-\infty} F_X(x) = 0$.
  ($F_X(\infty) = P(\Omega) = 1$. $F_X(-\infty) = P(\emptyset) = 0$.)
- $F_X(x)$ is right continuous, i.e., $\lim_{\epsilon\to 0,\ \epsilon>0} F_X(x+\epsilon) = F_X(x)$.
- $P(X = a) = F_X(a) - F_X(a^-)$ where
  $F_X(a^-) = \lim_{\epsilon\to 0,\ \epsilon>0} F_X(a-\epsilon) = P(X < a)$.
  ($\{X \le a\} = \{X < a\} \cup \{X = a\} \Rightarrow F_X(a) = F_X(a^-) + P(X = a)$.)
- For any Borel set A, $P(X \in A)$ can be determined from $F_X(x)$
- Notation: $X \sim F_X(x)$ means that $F_X(x)$ is the cdf of X.

Consider roll of a (6-sided) die with $\Omega = \{1, 2, 3, 4, 5, 6\}$. Suppose you bet \$10. that the outcome is 3. So you win \$10. if outcome $\omega = 3$, and lose \$10. if $\omega =$1, 2, 4, 5, or 6. Thus, the following function X defined on ω models the result of your bet:

$$X(\omega) = \begin{cases} 10 & \text{if } \omega = 3 \\ -10 & \text{otherwise} \end{cases}$$

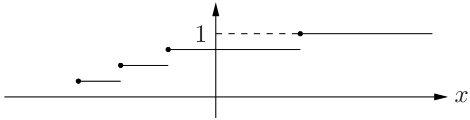
Find the cdf $F_X(x)$ assuming a fair die.
Assuming a fair die, $P(X = 10) = 1/6$ and $P(X = -10) = 5/6$. We have

$$\begin{aligned} x < -10 &: F_X(x) = P(X \le x) = 0 \\ x = -10 &: F_X(-10) = P(X < -10) + P(X = -10) = 0 + \frac{5}{6} = \frac{5}{6} \\ -10 < x < 10 &: F_X(x) = P(X \le -10) + P(-10 < X \le x) = \frac{5}{6} + 0 = \frac{5}{6} \\ x = 10 &: F_X(10) = P(X < 10) + P(X = 10) = \frac{5}{6} + \frac{1}{6} = 1 \\ x > 10 &: F_X(x) = P(X \le x) = F_X(10) + P(10 < X \le x) = 1 + 0 = 1 \end{aligned}$$


$F_X(x)$
1
$\frac{5}{6}$
0
x
−10
10

Probability Mass Function (PMF)

- A random variable is said to be discrete if $F_X(x)$ consists only of steps over a countable set $\mathcal{X}$. Since $\mathcal{X}$ ic countable, we may write $\mathcal{X} = \{x_i,\ i = 1, 2, \cdots\}$



- Hence, a discrete random variable can be completely specified by the probability mass function (pmf)

$$P_X(x) = P(X = x) \text{ for every } x \in \mathcal{X}$$

  Clearly $P_X(x) \geq 0$ and $\sum_{x \in \mathcal{X}} P_X(x) = 1$

- Notation: $X \sim P_X(x)$ means that the discrete random variable X has pmf $P_X(x)$.

Commonly Used Discrete Random Variables

- Bernoulli: $X \sim \text{Bern}(p)$ for $0 \leq p \leq 1$ has the pmf $P_X(1) = p$ and $P_X(0) = 1 - p$.
- Geometric: $X \sim \text{Geom}(p)$ for $0 \leq p \leq 1$ has the pmf $P_X(k) = p(1-p)^{k-1}$, $k = 1, 2, \cdots$.
- Binomial: $X \sim \text{Binom}(n, p)$ for integer $n > 0$ and $0 \leq p \leq 1$ has the pmf

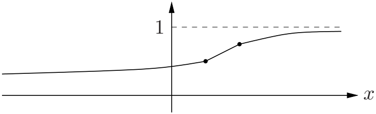
$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \cdots, n$$

- Poisson: $X \sim \text{Poisson}(\lambda)$ for $\lambda > 0$ has the pmf

$$P_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \cdots$$

Probability Density Function (pdf)

- A random variable X is said to be continuous if $F_X(x)$ is continuous



- If $F_X(x)$ is continuous and differentiable (except possibly over a countable set), then X can be completely specified by a probability density function (pdf) $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

- If $F_X(x)$ is differentiable everywhere, then (by definition)

$$\begin{aligned} f_X(x) &= \frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x} \end{aligned}$$

- Properties of pdf:
  - $f_X(x) \geq 0$ (since $f_X(x) = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x}$).
  - $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = P(X \leq x)$.
  - $\int_{-\infty}^{\infty} f_X(x) = 1 = F_X(\infty) - F_X(-\infty)$
  - For any event (Borel set) $A \subset \mathbb{R}$,

$$P(X \in A) = \int_{x \in A} f_X(x)\, dx$$

In particular,

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

- $f_X(x)$ should not be interpreted as the probability that $X = x$. In fact, $f_X(x)$ is not a probability measure since it can be > 1
- Notation: $X \sim f_X(x)$ means that X has pdf $f_X(x)$.

Using Delta Functions

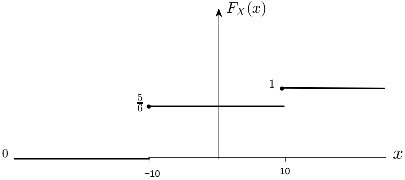
- Suppose $F_X(x)$ has a jump of height a at x_0. Obviously it is discontinuous at x_0, so its derivative does not exist. But we will use (Dirac) delta functions and define the pdf at $x = x_0$ as

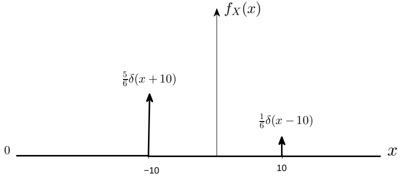
$$f_X(x_0) = a\,\delta(x - x_0)$$

- Recall the example (roll a die) with $\Omega = \{1, 2, 3, 4, 5, 6\}$ and

$$X(\omega) = \begin{cases} 10 & \text{if } \omega = 3 \\ -10 & \text{otherwise} \end{cases}$$

Its pdf is $f_X(x) = \frac{5}{6}\delta(x + 10) + \frac{1}{6}\delta(x - 10)$.





Classification of Random Variables

- Continuous random variable X: No delta functions in $f_X(x)$, equivalently, no jumps in $F_X(x)$.
- Discrete random variable X: Only delta functions in $f_X(x)$, equivalently, $F_X(x)$ is a staircase function. In this case, it is best to use pmf.
- Mixed random variable X: It is a mixture of the above two: some smooth nonzero parts in $f_X(x)$ together with some delta functions.

For any function $h(t)$ which has no delta functions,

$$\int_{T_1}^{T_2} h(t)\delta(t-t_0)\,dt = \begin{cases} h(t_0) & \text{if} \quad T_1 < t_0 < T_2 \\ 0 & \text{if} \quad t_0 < T_1 \text{ or } t_0 > T_2 \\ ? & \text{if} \quad t_0 = T_1 \text{ or } t_0 = T_2 \end{cases}$$

Suppose X has pdf $f_X(x) = 0.5\delta(x-5) + 0.05[u(x) - u(x-10)]$ Here $u(x)$ is the unit step function, i.e., $u(x) = 1$ for $x \geq 0$, and $u(x) = 0$ for $x < 0$. Find

- $P(X < 5) = \int_{-\infty}^{5^-} f_X(x)\,dx = 0.05\int_0^{5^-} dx + 0.5\int_0^{5^-} \delta(x-5)\,dx$
  $= 0.05\, x\Big|_0^5 + 0 = 0.25$
- $P(X \leq 5) = 0.25 + 0.5\int_0^{5^+} \delta(x-5)\,dx = 0.25 + 0.5 = 0.75$
- $P(X = 5) = \int_{5^-}^{5^+} f_X(x)\,dx = 0.5$
- $P(X = 2) = \int_{2^-}^{2^+} f_X(x)\,dx = 0$

Commonly Used Continuous Random Variables

- Uniform: $X \sim \mathrm{U}[a, b]$ for $a < b$ has the pdf

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- Exponential: $X \sim \mathrm{Exp}(\lambda)$ for $\lambda > 0$ has the pdf

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

- Laplace: $X \sim \mathrm{Laplace}(\lambda)$ for $\lambda > 0$ has the pdf

$$f_X(x) = \frac{\lambda}{2}\, e^{-\lambda\,|x|}\,,\ -\infty < x < \infty$$

- Gaussian: $X \sim \mathcal{N}(\mu, \sigma^2)$ with parameters μ and σ^2 has the pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,,\ -\infty < x < \infty$$

Expectation

- Let $X \in \mathcal{X}$ be a discrete r.v. with pmf $P_X(x)$ and let $g(x)$ be a function of x. The expectation (or expected value or mean) of $g(X)$ is defined as
$$E\{g(X)\} = \sum_{x \in \mathcal{X}} g(x) P_X(x)$$
- For a continuous/mixed r.v. $X \sim f_X(x)$, the expected value of $g(X)$ is
$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$
- Expectation is a linear operation, i.e., for any constants a and b
$$E\{a g_1(X) + b g_2(X)\} = a E\{g_1(X)\} + b E\{g_2(X)\}$$
In particular, $E\{a\} = a$.

- Remark: We know that a random variable is completely specified by its cdf (pdf, pmf), so why do we need expectation?
  - Expectation provides a summary or an estimate of the r.v. – a single number – instead of specifying the entire distribution
  - It is far easier to estimate the expectation of an r.v. from data than to estimate its distribution
  - Expectation can be used to bound or estimate probabilities of interesting events (as we shall see).

Mean and Variance

- The first moment (or mean) of $X \sim f_X(x)$ is

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

- The second moment (or mean squared or average power) of X is

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx$$

- The variance of X is

$$\mathrm{Var}(X) = E\{(X - E\{X\})^2\} = E\{X^2\} - (E\{X\})^2$$

  It then follows that $E\{X^2\} \geq (E\{X\})^2$.
- The standard deviation of X is defined as $\sigma_X = \sqrt{\mathrm{Var}(X)}$, or $\sigma_X^2 = \mathrm{Var}(X)$
- In general, the kth moment (k a positive integer) of X is

$$E\{X^k\} = \int_{-\infty}^{\infty} x^k f_X(x)\, dx$$

Mean and Variance of Some Common RVs

Random Variable	Mean	Variance
$\mathrm{Bern}(p)$	p	$p(1-p)$
$\mathrm{Geom}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathrm{Binom}(n,p)$	np	$np(1-p)$
$\mathrm{Poisson}(\lambda)$	λ	λ
$\mathrm{U}[\,a,b\,]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathrm{Exp}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathrm{Laplace}(\lambda)$	0	$\frac{2}{\lambda^2}$
$\mathcal{N}\left(\mu,\sigma^2\right)$	μ	σ^2

Conditional Distribution

- Conditional CDF Conditional CDF $F_{X|B}(x|B)$ of random variable X given some event (Borel set in $\mathbb{R}$) B (involving X) with $P(B) \neq 0$, is defined as

$$F_{X|B}(x|B) \stackrel{\text{def}}{=} P(X \leq x|B) = \frac{P(\{X \leq x\} \cap B)}{P(B)}$$

- Conditional pdf Conditional pdf $f_{X|B}(x|B)$ of random variable X given some event B (involving X) with $P(B) \neq 0$, is defined as

$$f_{X|B}(x|B) \stackrel{\text{def}}{=} \frac{d\,F_{X|B}(x|B)}{dx}.$$

- Conditional pmf is defined similarly.
- Conditional CDF, pdf and pmf satisfy all properties of CDF, pdf and pmf, respectively. For instance, $F_{X|B}(x|B)$ is a non-decreasing function of x, $F_{X|B}(-\infty|B) = 0$, $F_{X|B}(\infty|B) = 1$, $\int_{-\infty}^{\infty} f_{X|B}(x|B)\,dx = 1$.

Suppose $X \sim \exp(0.1)$ where X represents time-to-failure of some system/component in years What is the (conditional) average lifetime of the system given that the system has survived for 5 years?

We have $f_X(x) = 0.1e^{-0.1x}$ for $x \geq 0$. Then $E\{X\} = 1/0.1 = 10$ years, i.e., (unconditional) average life of the system is 10 years. If $X \sim \exp(\lambda)$, then using integration by parts,

$$E\{X\} = \int_0^\infty x\lambda e^{-\lambda x}\,dx = x\lambda \frac{e^{-\lambda x}}{-\lambda}\Big|_0^\infty - \int_0^\infty \lambda \frac{e^{-\lambda x}}{-\lambda}\,dx$$
$$= 0 + \int_0^\infty e^{-\lambda x}\,dx = \frac{e^{-\lambda x}}{-\lambda}\Big|_0^\infty = \frac{1}{\lambda}.$$

We need to compute $E\{X|B\} = \int_0^\infty x f_{X|B}(x|B)\,dx$ where $B = \{X > 5\}$.

We need to determine $F_{X|B}(x|B)$ first. We have

$$F_{X|B}(x|B) = \frac{P(\{X \leq x\} \cap \{X > 5\})}{P(X > 5)} = \begin{cases} 0 & \text{if } x \leq 5 \\ \frac{P(5 < X \leq x)}{e^{-5\lambda}} & \text{if } x > 5 \end{cases}$$

Thus

$$F_{X|B}(x|B) = \begin{cases} 0 & \text{if } x \leq 5 \\ \frac{e^{-5\lambda} - e^{-\lambda x}}{e^{-5\lambda}} & \text{if } x > 5 \end{cases}$$

$$\Rightarrow f_{X|B}(x|B) = \begin{cases} 0 & \text{if } x \leq 5 \\ \frac{\lambda e^{-\lambda x}}{e^{-5\lambda}} = \lambda e^{-\lambda(x-5)} & \text{if } x > 5 \end{cases}$$

Hence, $E\{X|X > 5\} = \int_5^\infty x\lambda e^{-\lambda(x-5)}\, dx = \int_0^\infty (y+5)\lambda e^{-\lambda y}\, dy$. This leads to $E\{X|X > 5\} = \frac{1}{\lambda} + 5 = 10 + 5 = 15$ years. That is, having survived for 5 years, the remaining average life of the system is still 10 years! The exponential distribution is therefore called memoryless.

Some Useful Formulas: Conditional Distribution

- Given a partition $\{A_1, A_2, \cdots, A_n\}$ of Ω,

$$F_X(x) = \sum_{i=1}^{n} F_{X|A_i}(x|A_i)P(A_i), \quad f_X(x) = \sum_{i=1}^{n} f_{X|A_i}(x|A_i)P(A_i)$$

- Bayes' Formula for Conditional pdf: If $P(B) \neq 0$, what is $P(B|X = x)$? If X is a continuous random variable, then $P(X = x) = 0$. Consider

$$P(B|x < X \leq x + \Delta) = \frac{P(B \cap \{x < X \leq x + \Delta\})}{P(\{x < X \leq x + \Delta\})}$$

$$= \frac{P(\{x < X \leq x + \Delta\}|B)P(B)}{P(\{x < X \leq x + \Delta\})}$$

$$= \frac{[F_{X|B}(x + \Delta|B) - F_{X|B}(x|B)]P(B)}{F_X(x + \Delta) - F_X(x)} = \frac{\frac{F_{X|B}(x+\Delta|B) - F_{X|B}(x|B)}{\Delta} P(B)}{\frac{F_X(x+\Delta) - F_X(x)}{\Delta}}$$

- Now let $\Delta \to 0$:

$$\lim_{\Delta \to 0} P(B|x < X \le x + \Delta) = \frac{f_{X|B}(x|B)P(B)}{f_X(x)} = P(B|X = x)$$

- Cross multiply and integrate:

$$P(B) \int_{-\infty}^{\infty} f_{X|B}(x|B)\,dx = \int_{-\infty}^{\infty} P(B|X = x) f_X(x)\,dx$$

$$\Rightarrow P(B) = \int_{-\infty}^{\infty} P(B|X = x) f_X(x)\,dx$$

A box contains 20 lightbulbs with average lifetime of 1000 hours and 30 lightbulbs with average lifetime of 800 hours. The time-to-failure of each of these lightbulbs follows an exponential distribution. A lightbulb is selected at random from the box. The lightbulbs are not labeled with their expected lifetime. Determine the pdf of the time-to-failure of the selected bulb. What is the average lifetime of the the selected bulb?

Let T, T_A and T_B denote the expected lifetime of the selected bulb, bulbs with lifetime of 1000 hours and bulbs with lifetime of 800 hours, respectively, all measured in hours. We are given $T_A \sim \exp(1/1000)$ and $T_B \sim \exp(1/800)$, since if $X \sim \exp(\lambda)$, then $E\{X\} = \lambda^{-1}$. Denote the two sets of bulbs as A and B with lifetimes T_A and T_B, respectively. When a bulb is selected at random from the box, the probability $P(A)$ that it is of type A, is $P(A) = 20/(20+30) = 0.4$. Then $P(B) = 1 - 0.4 = 0.6$. By law of total probability,

$$f_T(t) = f_{T|A}(t|A)P(A) + f_{T|B}(t|B)P(B) = 0.4\lambda_1 e^{-\lambda_1 t} + 0.6\lambda_2 e^{-\lambda_2 t}$$

where $\lambda_1 = 1/1000$ and $\lambda_2 = 1/800$. The expected lifetime is

$$E\{T\} = 0.4E\{T_A\} + 0.6E\{T_B\} = 0.4 \times 1000 + 0.6 \times 800 = 880 \text{ hours}$$