Random Variables

Johannes Wissel

School of Operations Research & Information Engineering
Cornell University

Spring 2024

Reading: Devore 2.5, 3.1–3.3

Random Variables

Usually we're interested in *numerical* characteristics of the outcome of a random phenomenon.

Example: For n flips of a coin, the sample space is

$$\mathcal{S} = \{(x_1, \ldots, x_n) : x_i \in \{\mathsf{H}, \mathsf{T}\}, i = 1, \ldots, n\}.$$

We may only be interested in the number N of heads obtained from n coin flips.

Random Variables

Engineer's Definition: a *random variable* is a number whose value is unknown.

Mathematician's Definition: a random variable is a function from the sample space $\mathcal{S}$ to $\mathbb{R}$.

Example: The number N of heads obtained from 3 coin flips is a random variable, e.g.,

$$N((\mathrm{H},\ \mathrm{T},\ \mathrm{H})) = 2$$

$$N((\mathrm{T},\ \mathrm{T},\ \mathrm{T})) = 0$$

Example: Quality Control

The number of flaws on each of 20 silicon wafers is recorded. The sample space is

$$\mathcal{S} = \{(x_1, \ldots, x_{20}) : x_i \in \{0, 1, 2, \ldots\}, i = 1, \ldots, 20\}$$

The following are random variables:

- T = total number of flaws:

$$T((x_1, \ldots, x_{20})) = \sum_{i=1}^{20} x_i$$

- M = maximum number of flaws on any wafer:

$$M((x_1, \ldots, x_{20})) = \max\{x_1, \ldots, x_{20}\}$$

Random Variables: Notation

Distinguish between random variables and their observed values.

Example: (See previous slide.) If it's observed that each wafer has 1 flaw, then the observed value of the random variable T is $t = 20$.

In general, we use

- upper-case letters (e.g., X) for random variables;
- lower-case letters (e.g., x) for the particular values they can take.

Types of Random Variables

Discrete: The set of possible values that the random variable can take is finite (e.g., $\{\text{H}, \text{T}\}$) or countably infinite (e.g., $\{1, 2, \dots\}$).

Continuous: The set of possible values that the random variable can take is an interval (e.g., $[0, 10]$, $[0, \infty)$ or $(-\infty, \infty)$).

Example:

- The random variables T and M on slide 3 are discrete.
- The failure time of a battery can be modeled with a continuous random variable (possible values $= [0, \infty)$).

Probability Mass Functions

Example: Sample, at random, a family that has 3 children and note their genders (f for female, m for male) in the order of their birth. The sample space is

$$\mathcal{S} = \{fff, ffm, fmf, mff, mmf, mfm, fmm, mmm\}.$$

Consider the random variable $N =$ number of girls.

Question: What is $P(N = 2)$? (What're you assuming?)

A. $1/8$
B. $2/8$
C. $3/8$
D. $5/8$

Probability Mass Functions

For each discrete random variable X defined on the sample space $\mathcal{S}$, the underlying probability model $(\mathcal{S}, \mathcal{E}, P)$ induces a "probability mass function" $p_X(x)$ on the set of possible values of X.

Example: For the example on slide 6, when $p_s = 1/8$ for every $s \in \mathcal{S} = \{fff, ffm, fmf, mff, mmf, mfm, fmm, mmm\}$, the pmf $p_N(n)$ for N is

$$p_N(0) = \frac{1}{8},$$
$$p_N(1) = \frac{3}{8},$$
$$p_N(2) = \frac{3}{8},$$
$$p_N(3) = \frac{1}{8}.$$

Probability Mass Functions

Definition: For a discrete random variable X defined for a probability model $(\mathcal{S}, \mathcal{E}, P)$, the *probability mass function (PMF)* of X is

$$p_X(x) = P(X = x) = P(\{s \in \mathcal{S} : X(s) = x\})$$

Some properties:

$$p_X(x) = 0 \qquad \text{if } X \text{ never takes the value } x$$

$$p_X(x) \geq 0 \qquad \text{for all possible values } x \text{ of } X$$

$$\sum_{\text{all possible } x} p_X(x) = 1$$

Example: Geometric Random Variables

Flip a coin, that comes up heads with probability $\rho \in [0, 1]$, until it comes up heads.

$$\mathcal{S} = \{\mathsf{H}, \mathsf{TH}, \mathsf{TTH}, \mathsf{TTTH}, \dots\}$$

The random variable $X : \mathcal{S} \to \{1, 2, \dots\}$ that gives the number of flips needed to get the first heads is a *geometric random variable*.

PMF of X: For $x = 1, 2, \dots,$

$$\begin{aligned} p_X(1) &= \rho \\ p_X(2) &= (1-\rho)\rho \\ p_X(3) &= (1-\rho)^2\rho \\ &\vdots \\ p_X(x) &= (1-\rho)^{x-1}\rho \end{aligned}$$

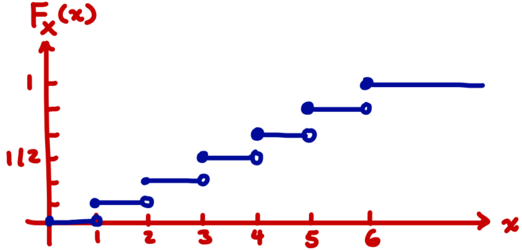
Cumulative Distribution Functions

For a random variable, the underlying probability model also induces a **cumulative distribution function (CDF)** F_X where, for $x \in \mathbb{R}$,

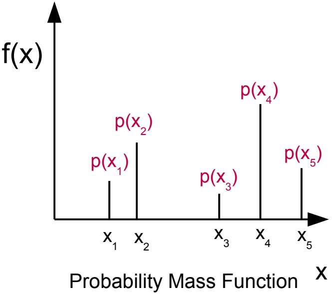
$$F_X(x) = P(X \leq x) = \sum_{y \leq x} p_X(y)$$

Example: Throwing a fair 6-sided die:

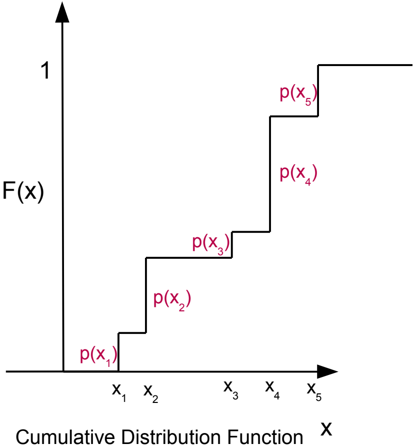
For X= result of the toss :



PMFs vs. CDFs



Probability Mass Function



Cumulative Distribution Function

Cumulative Distribution Functions

You're given the PMF p_X and CDF F_X of a random variable X.

Questions:

Which of the following equals $P(X > x)$?

A. $p_X(x)$

B. $1 - p_X(x)$

C. $F_X(x)$

D. $1 - F_X(x)$

For $a < b$, which of the following equals $P(a < X \leq b)$?

A. $F_X(b) - F_X(a)$

B. $F_X(a) - F_X(b)$

C. $F_X(a) + F_X(b)$

D. $F_X(b)/F_X(a)$

Expectation

Suppose we have a random experiment/phenomenon modeled with the probability model $(\mathcal{S}, \mathcal{E}, P)$.

We're interested in a random variable X on $(\mathcal{S}, \mathcal{E}, P)$.

We:

1. Observe the experiment/phenomenon.
2. Get an outcome $s \in \mathcal{S}$.
3. Get a value $X(s)$ of the random variable.

Question: What value of the random variable do we *expect* to get before we perform the observation?

Expectation: Long-Run Average Interpretation

Let $\mathcal{X}$ be the set of all possible values of the random variable X.

Observe the value of X a large number (say, M) of times:

$$x_1, \ldots, x_M$$

Take their average:

$$\frac{x_1 + \cdots + x_M}{M} = \sum_{x \in \mathcal{X}} x \cdot \frac{(\#\ x_i\text{'s} = x)}{M} \approx \sum_{x \in \mathcal{X}} x \cdot p_X(x).$$

Expectation

Definition: The *expectation* or *mean* $E(X)$ of a discrete r.v. X is

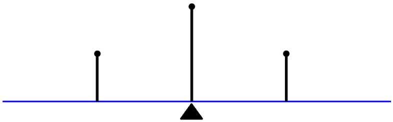
$$E(X) = \sum_{x \in \mathcal{X}} x \cdot P(X = x) = \sum_{x \in \mathcal{X}} x \cdot p_X(x).$$

In words:

1. For each possible value of x, compute $x \cdot P(X = x)$.
2. Sum the values in step 1.

Expectation: Physical Interpretation

$E(X)$ is the "balance point" of the r.v. X's PMF.



Note:

- $E(X)$ need not be a possible value of X.
- $E(X)$ can be infinity (e.g., when $p_X(x) = (6/\pi^2)/x^2$ for $x = 1, 2, \ldots$).

Expectation

Example: Consider throwing a fair die, where $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$, and define the r.v. X as follows:

s	1	2	3	4	5	6
$P(\{s\})$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$
$X(s)$	0	0	1	1	2	2

Then $E(X)$ is

$$\begin{aligned} E(X) &= \sum_{x \in \mathcal{X}} x \cdot P(X = x) \\ &= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) \\ &= 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} \\ &= 1. \end{aligned}$$

Example: Geometric Random Variable (Rephrased)

Bernoulli Trial: random experiment that's either a "success" (denoted S) or "failure" (denoted F).

Consider performing a sequence of Bernoulli trials, each having success probability $\rho \in [0, 1]$, until the first "success". Define the r.v. X by:

s	S	FS	FFS	$FFFS$	$\cdots$
$P(\{s\})$	ρ	$(1-\rho)\rho$	$(1-\rho)^2\rho$	$(1-\rho)^3\rho$	$\cdots$
$X(s)$	1	2	3	4	$\cdots$

X is a *geometric random variable*.

Example: Geometric Random Variable

The mean of a geometric random variable X is

$$
\begin{aligned}
E(X) &= \sum_{x \in \mathcal{X}} x \cdot p_X(x) \\
&= \sum_{x=1}^{\infty} x(1-\rho)^{x-1}\rho \\
&= \rho \sum_{x=1}^{\infty} x(1-\rho)^{x-1} \\
&= \rho \frac{\mathrm{d}}{\mathrm{d}q}\left[\sum_{x=1}^{\infty} q^x\right] \\
&= \rho \frac{\mathrm{d}}{\mathrm{d}q}\left[\frac{1}{1-q} - 1\right] \\
&= \rho \frac{1}{(1-q)^2} = \frac{\rho}{\rho^2} = \frac{1}{\rho}.
\end{aligned}
$$

Let's set $q = 1 - \rho$

Expectation

Let the value of the random variable Y be 1 if the number of Bernoulli trials (with "success" probability ρ) required to get the first "success" is at least 2, and let the value of Y be 0 otherwise.

Question: What is $E(Y)$?

A. ρ
B. $1 - \rho$
C. $\rho/(1 - \rho)$
D. $(1 - \rho)\rho + (1 - \rho)^3\rho + (1 - \rho)^5 + \cdots$

$$\begin{aligned} E(Y) &= 1 \cdot P(\geq 2 \text{ trials to get success}) + 0 \cdot P(\text{otherwise}) \\ &= 1 - P(1 \text{ trial to get success}) = 1 - \rho. \end{aligned}$$

Expectation of Functions of Random Variables

Consider a random variable X and a function $h : \mathbb{R} \to \mathbb{R}$.

Then $h(X)$ is also a random variable, and

$$E(h(X)) = \sum_{x \in \mathcal{X}} h(x)P(X = x) = \sum_{x \in \mathcal{X}} h(x)p_X(x).$$

Example: What is $E(X^2)$ where X is the number obtained from a fair die roll?

$$1^2(1/6) + 2^2(1/6) + 3^2(1/6) + \cdots + 6^2(1/6) = 15.2$$

Properties of Expectation

Expectation of a Constant: For a constant c,

$$E(c) = c$$

Linearity: If X and Y are r.v.'s and a, b are constants, then

$$E(aX + bY) = aE(X) + bE(Y)$$

In particular, if $h(x) = ax + b$ is a linear function, then

$$E(h(X)) = h(E(X)) = aE(X) + b$$

Properties of Expectation

In general, $E(h(X)) \neq h(E(X))$:

Flip a fair coin twice.

$X = \#$ of heads, $h(x) = 2x^2$.

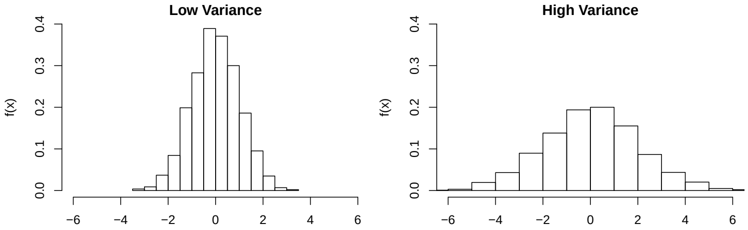
$$\begin{aligned} E(h(X)) &= \sum_{x \in \mathcal{X}} h(x) P(X = x) \\ &= 2(0)^2 P(X = 0) + 2(1)^2 P(X = 1) + 2(2)^2 P(X = 2) \\ &= 2 \cdot \frac{1}{2} + 8 \cdot \frac{1}{4} = 3. \end{aligned}$$

$$\begin{aligned} h(E(X)) &= 2\Big(0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2)\Big)^2 \\ &= 2\left(\frac{1}{2} + 2 \cdot \frac{1}{4}\right)^2 = 2(1)^2 = 2. \end{aligned}$$

Variance

Expectation is a measure of where a r.v.'s values are located.

The variance of a r.v. X measures how spread out X's possible values are around $\mu = E(X)$.



Variance

Definition: The *variance* of a r.v. X is

$$\sigma_X^2 = \mathrm{Var}(X) = E((X-\mu)^2) = \sum_{x\in\mathcal{X}} (x-\mu)^2 p_X(x),$$

and the *standard deviation* of X is

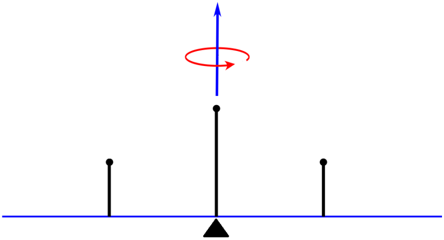
$$\sigma_X = \sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

Question: Why not measure how spread out X is by using

$$\sum_{x\in\mathcal{X}} (x-\mu) p_X(x) \text{ ?}$$

Variance: Physical Interpretation

σ_X^2 measures how hard it is to spin the PMF of X (i.e., the PMF's *"moment of inertia"*) around its balance point $E(X)$ (i.e., the PMF's "center of mass").



More spread out $\implies$ bigger variance.

Example: Games

Consider the following 3 games:

Game 1: Flip a fair coin; win \$1 if heads, and lose \$1 if tails.

Game 2: Roll two fair six-sided die; win \$1 if the result is not 12, and lose \$35 otherwise.

Game 3: Same as Game 2, but win \$100 and lose \$3500 instead.

Question: Which game would you play?

A. Game 1. (expectation $= 0$, variance $= 1$)

B. Game 2. (expectation $= 0$, variance $= 35$)

C. Game 3. (expectation $= 0$, variance $= 350,000$)

Properties of Variance

If a and b are constants, then

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

$$\begin{aligned}\mathrm{Var}(aX+b) &= E\Big(\big(aX+b-E(aX+b)\big)^2\Big)\\ &= E\Big(\big(aX-aE(X)\big)^2\Big)\\ &= a^2E\Big(\big(X-E(X)\big)^2\Big) = a^2\mathrm{Var}(X).\end{aligned}$$

If $P(X=c)=1$ for some constant c, then

$$\mathrm{Var}(X) = 0.$$

$$\mathrm{Var}(X) = E\Big(\big(X-E(X)\big)^2\Big) = (c-c)^2 = 0.$$

Alternative Variance Formula

Often good for computing variance by hand (but not on a computer):

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

$$\begin{aligned}
\mathrm{Var}(X) &= E\Big(\big(X - E(X)\big)^2\Big) \\
&= E\Big(X^2 - 2X \cdot E(X) + E(X)^2\Big) \\
&= E(X^2) - 2E(X) \cdot E(X) + E(X)^2 \\
&= E(X^2) - E(X)^2.
\end{aligned}$$

Example: Bernoulli R.V.

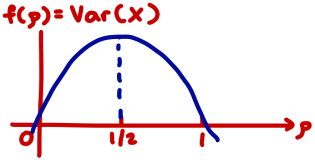
Given $\rho \in [0, 1]$, a *Bernoulli random variable* X is defined as follows:

- Possible values: $\mathcal{X} = \{0, 1\}$.
- $p_X(0) = 1 - \rho$ and $p_X(1) = \rho$.

Question: For what value of ρ is $\mathrm{Var}(X)$ maximized?

$E(X) = p$, $E(X^2) = p$

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



$\Rightarrow p = \frac{1}{2}$ maximizes $\mathrm{Var}(X)$

Standardizing a Random Variable

Suppose X is a r.v. with expectation μ and standard deviation σ. The *standardized* version of X is

$$X^* = \frac{X - \mu}{\sigma}.$$

Then X^* has expectation 0 and variance 1:

$$E(X^*) = \frac{1}{\sigma}(E(X) - \mu) = 0$$

$$\mathrm{Var}(X^*) = \frac{1}{\sigma^2}\mathrm{Var}(X) = 1$$

Summary

- A *random variable (r.v.)* is a real-valued function on the sample space $\mathcal{S}$.
- A discrete r.v. is characterized by its *probability mass function (PMF)*, or its *cumulative distribution function (CDF)*.
- The mean of a r.v. X is the "center of mass", and the variance of X is the "moment of inertia", of the PMF of X.