Random Variables and Probability Distributions

B39AX — Fall 2023

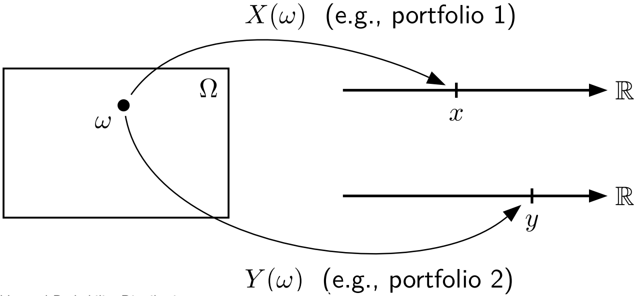
Heriot-Watt University

Random variables

Often *indirect* outcomes of an experiment are more interesting than *direct* outcomes.

Example: profits in the stock market (indirect) vs stock values (direct)

Random variable: a function from sample space Ω to $\mathbb{R}$



Random variables

Unfortunate term: random variables (RVs) are functions, not variables

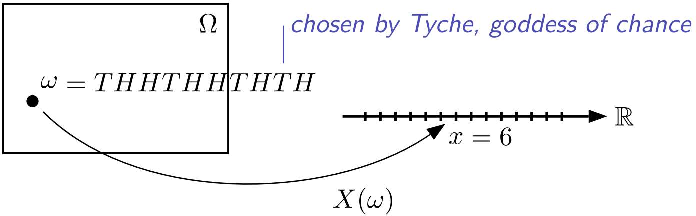
Convention: write RVs in upper case, numbers in its range in lower case

Example: Let RV X represent total number of heads in 10 flips of a coin

$$\mathbb{P}(X = x) = \binom{10}{x} \cdot 0.5^x \cdot 0.5^{10-x}$$

binomial distribution

upper case (RV) **lower case** (concrete number)



Random variables

Another Example:

Consider the value of 20 stocks $\Omega = \big(\mathbb{R}_+, \mathbb{R}_+, \ldots, \mathbb{R}_+\big)$, e.g.,

(Open AI, Amazon, Meta, Google, $\ldots$)

I invested $\pounds 5000$ in 1 Open AI, 30 Amazon, and 5 Meta stocks

My profit/loss on day t: $\quad X_t(\omega) = \omega_1 + 30\,\omega_2 + 5\,\omega_3 - 5000$

When day t is over, my profit/loss on day t is $x_t = X_t(\omega^{\text{act}})$.

Tyche

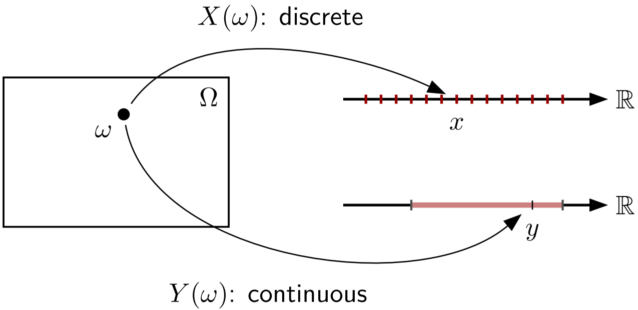
Discrete and continuous RVs

Discrete random variable

It takes values in a countable subset $\{x_1, x_2, \dots\}$ of $\mathbb{R}$

Continuous random variable

It takes values in a continuous (uncountable) subset of $\mathbb{R}$



Discrete and continuous RVs

Examples:

- X : total number of heads in 10 flips of a coin

  $X : \Omega \to \{0, 1, \ldots, 10\}$ ***discrete***

  $\{TT \cdots T\,, TT \cdots H\,, \ldots\,, HH \cdots H\}$

- Y : length of the tallest tree on campus

  $Y : \Omega \to \mathbb{R}_{++} = \{x : x > 0\}$ ***continuous***

But how do we characterize/describe a random variable?

Exercise

Consider the random experiment of tossing 2 dice independently.

$$\Omega = \Big\{(1,1),(1,2),\ldots,(2,1),\ldots,(6,5),(6,6)\Big\}$$ ***36 events***

We are interested not in the values of the dice, but only in their sum.

Define the random variable (RV) X as *the sum of both dice.*

Compute $\mathbb{P}(X = x)$, for $x = 2, 3, \ldots, 12$.

Ans: $\mathbb{P}(X = x) = (6 - |x - 7|)/36$

PMF, PDF, and CDF

Probability mass function (pmf) of a ***discrete*** RV X:

$$p_X(x) := \mathbb{P}\big(X = x\big)\,, \qquad \text{for } x \text{ in a countable set.}$$

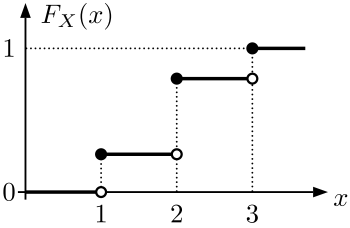
Probability density function (pdf) of a ***continuous*** RV X is f_X s.t.

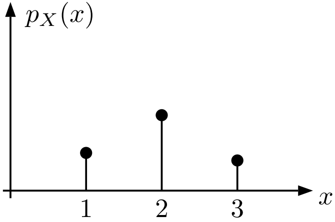
$$\mathbb{P}\big(a \le X \le b\big) = \int_a^b f_X(t)\,dt\,, \qquad \text{for any } a \le b.$$

Cumulative distribution function (cdf) of ***any*** RV X is the function

$$F_X(x) = \mathbb{P}\big(X \le x\big) = \begin{cases} \displaystyle\sum_{i \le x} p_X(i) & \text{, if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{x} f_X(t)\,dt & \text{, if } X \text{ is continuous.} \end{cases}$$

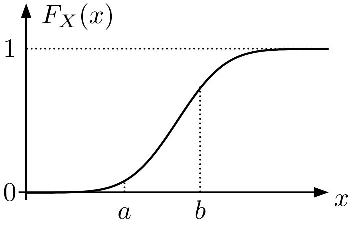
Discrete RV

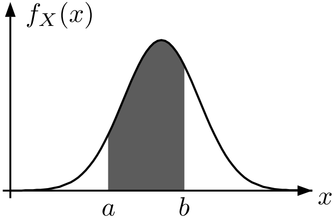




$$\mathbb{P}\big(X = x\big) = p_X(x)$$

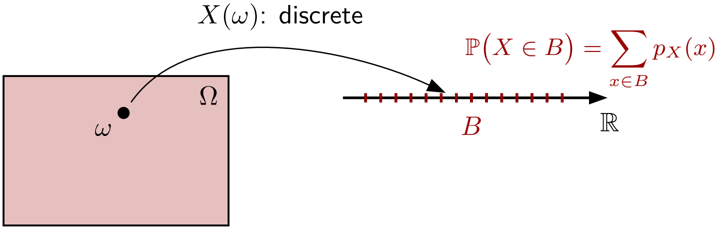
Continuous RV





$$\mathbb{P}\big(a \le X \le b\big) = F_X(b) - F_X(a)$$
$$= \int_a^b f_X(x)\,dx$$

Properties



Because $\mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\omega \in \Omega) = 1$,

- $\displaystyle\sum_{x=-\infty}^{+\infty} p_X(x) = 1$
- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$ (for continuous RVs)
- $F_X(\infty) = \mathbb{P}(X \leq +\infty) = 1$ (for both discrete and continuous RVs)

Example: Bernoulli random variable

X is a *Bernoulli random variable* if it takes only two values:

- 1 (success) with probability p, i.e., $p_X(1) = \mathbb{P}(X = 1) = p$
- 0 (failure) with probability $1 - p$, i.e., $p_X(0) = \mathbb{P}(X = 0) = 1 - p$

We also say that X has *Bernoulli distribution*, and write

$$\boxed{X \sim \mathsf{Ber}(p)}$$

It is a valid probability distribution, because

$$\sum_k p_X(k) = \sum_{k=0}^{1} p_X(k) = p_X(0) + p_X(1) = (1 - p) + p = 1$$

Example:

$X =$ "Randomly selected student scores ≥ 90 in exam", $\quad X \sim \mathsf{Ber}(0.1)$

Joint PMF, PDF, and CDF

Joint pmf of two ***discrete*** RVs X and Y is the function

$$p_{XY}(x\,,\,y) := \mathbb{P}\big(X = x\,,\,Y = y\big) = \mathbb{P}\big(\{X = x\} \cap \{Y = y\}\big)$$

Joint pdf of a ***continuous*** RVs X and Y is $f_{XY}(x\,,\,y)$ such that

$$\mathbb{P}\big(a \leq X \leq b\,,\,c \leq Y \leq d\big) = \int_c^d \int_a^b f_{XY}(x,y)\,dx\,dy$$

Joint cdf of ***any*** RVs X and Y is the function

$$F_{XY}(x\,,\,y) = \mathbb{P}\big(X \leq x\,,\,Y \leq y\big)$$

Example

Let $X =$ "Sunny *today*", $Y =$ "Sunny *tomorrow*"

The joint PMF can be given in tabular form:

		Y: S	Y: NS	
X	S	$\frac{1}{100}$	$\frac{9}{100}$	$\boxed{\frac{1}{10}}$
	NS	$\frac{9}{100}$	$\frac{81}{100}$	$\boxed{\frac{9}{10}}$
		$\frac{1}{10}$	$\frac{9}{10}$	1

$\mathbb{P}(X = \mathsf{S}, Y = \mathsf{S}) + \mathbb{P}(X = \mathsf{S}, Y = \mathsf{NS}) = \mathbb{P}(X = \mathsf{S})$

$\mathbb{P}(X = \mathsf{NS}, Y = \mathsf{S}) + \mathbb{P}(X = \mathsf{NS}, Y = \mathsf{NS}) = \mathbb{P}(X = \mathsf{NS})$

Marginal PMFs:

$$p_X(x) = \sum_y p_{XY}(x, y) \qquad p_Y(y) = \sum_x p_{XY}(x, y)$$

Independence of random variables

The random variables X and Y are ***independent*** if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all x and y.

Equivalently, for all x and y,

- X and Y are independent if $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$
- X and Y (discrete) are independent if $p_{XY}(x, y) = p_X(x) \cdot p_Y(y)$
- X and Y (continuous) are independent if $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$

Are X and Y in the previous example independent?

Expected value

Expectation summarizes all possible outcomes of a RV into one number.

Examples

Expected returns from the stock market, expected # of students in class

Expected value of RV X is represented as $\mathbb{E}[X]$.

If X is discrete,

$$\mathbb{E}[X] = \sum_k k \cdot p_X(k)$$

If X is continuous,

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

Expected value

Example: Let $X \sim \mathsf{Ber}(p)$.

$$p_X(k) = \begin{cases} 1 - p & , \ k = 0 \\ p & , \ k = 1 \end{cases}$$

Then,

$$\mathbb{E}[X] = \sum_{k=0}^{1} k \cdot p_X(k) = 0 \cdot (1 - p) + 1 \cdot p = p$$

Expected value of a function

$$\begin{cases} X \text{ is a RV} \\ \\ g : \mathbb{R} \to \mathbb{R} \text{ is a generic function} \end{cases} \implies \qquad g(X) \; : \text{ is also a RV}$$

Expected value of $g(X)$:

- If X is discrete and has pmf $p_X(k)$,

$$\mathbb{E}\big[g(X)\big] = \sum_{k=-\infty}^{+\infty} g(k) \cdot p_X(k)$$

- If X is continuous and has pdf $f_X(x)$,

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

Properties of the expected value

- The expected value is a linear operator: for any $a, b \in \mathbb{R}$,

$$\mathbb{E}\big[a\,X + b\big] = a\,\mathbb{E}[X] + b$$

- For any RVs X and Y,

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

- If X and Y are independent,

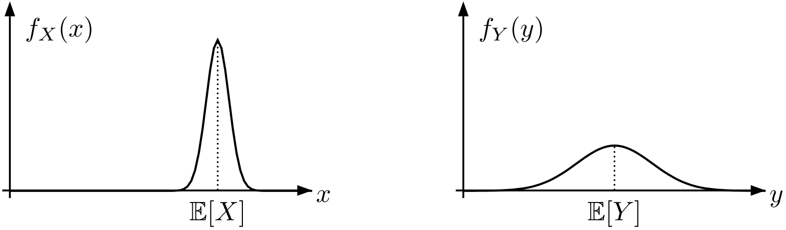
$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Proofs at the end of the slides

Variance

Variance of RV $\boldsymbol{X}$: $\quad \mathsf{Var}(X) = \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big]$

$\mathsf{Var}(X)$ measures the dispersion of the distribution of X around $\mathbb{E}[X]$:



$$\mathsf{Var}(X) < \mathsf{Var}(Y)$$

Standard deviation of RV $\boldsymbol{X}$: $\quad \sigma_X = \sqrt{\mathsf{Var}(X)} \quad$ (same units as X)

Properties of the variance

- $\mathsf{Var}(X) \geq 0$ for any RV X
- $\mathsf{Var}(X) = \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \mathbb{E}\big[X^2\big] - \Big(\mathbb{E}[X]\Big)^2$
- $\mathsf{Var}(a\,X + b) = a^2\,\mathsf{Var}(X)$, for any $a, b \in \mathbb{R}$
- If X and Y are independent, $\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$

Proofs at the end of the slides

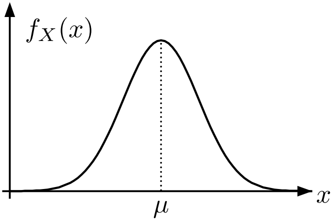
The normal distribution

X has ***Gaussian*** or ***normal distribution***, $X \sim \mathcal{N}(\mu, \sigma^2)$, if its pdf is

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

It can be shown that

- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
- $\mathbb{E}[X] = \mu$
- $\mathsf{Var}(X) = \sigma^2$



X has ***standard normal distribution*** if $X \sim \mathcal{N}(0, 1)$; its cdf is $\Phi(z)$

$X \sim \mathcal{N}(0, 1)$ arises "everywhere", because of the central limit theorem

The normal distribution

Normality is preserved under linear transformations

If $X \sim \mathcal{N}(\mu,\, \sigma^2)$ and $a \neq 0$ and b are scalars, then

$$Y = a\,X + b$$

is also normal with

$$\mathbb{E}[Y] = a\,\mu + b \qquad\qquad \mathsf{Var}(Y) = a^2\sigma^2$$

That is,

$$Y \sim \mathcal{N}\Big(a\,\mu + b,\; a^2\sigma^2\Big)$$

The normal distribution

Let $X \sim \mathcal{N}(\mu, \sigma^2)$

$\mathbb{P}(a \leq X \leq b)$ can be computed using tables for $Z \sim \mathcal{N}(0,1)$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830

The values of this table give $\Phi(z) := \mathbb{P}(Z \leq z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$

For example, $\Phi(0.42) = 0.6628$

The normal distribution

Procedure for computing $\mathbb{P}(a \leq X \leq b)$ for $X \sim \mathcal{N}(\mu, \sigma^2)$

- Transform X into $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$
- Use symmetry of Z around the origin to compute $\Phi((a-\mu)/\sigma)$ and $\Phi((b-\mu)/\sigma)$ from the tables (draw the pdf of Z)
- $\mathbb{P}(a \leq X \leq b) = \Phi((b-\mu)/\sigma) - \Phi((a-\mu)/\sigma)$

Justification:

$$\begin{aligned}
\mathbb{P}(a \leq X \leq b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)
\end{aligned}$$

Exercise

The number new students in EEE at HW each year is modeled as a normal RV with mean 60 and standard deviation 20. What is the probability that next year we will have more than 80 new students?

Ans: $\simeq 0.1587$

Other important distributions

Binomial distribution

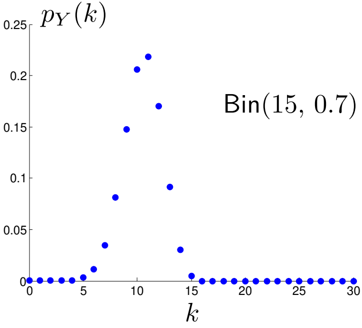
Perform n independent Bernoulli trials $X_1, \ldots, X_n$, where $X_i \sim \text{Ber}(p)$

Total # of successes $Y = X_1 + \cdots + X_n$ has *binomial distribution*

$Y \sim \text{Bin}(n, p)$

It can be shown that

- $p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k}$
  $k = 0, 1, \ldots, n$
- $\sum_{k=0}^{n} p_Y(k) = 1$
- $\mathbb{E}[Y] = np$
- $\text{Var}(Y) = np(1-p)$



Other important distributions

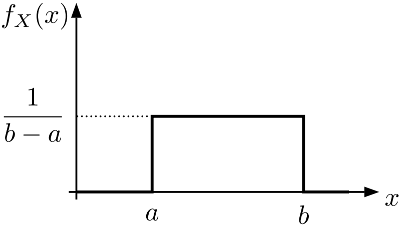
Continuous uniform distribution

X has *continuous uniform distribution* in $[a, b]$, $X \sim \mathcal{U}(a, b)$, if

$$f_X(x) = \begin{cases} \frac{1}{b-a} & , \; x \in [a, b] \\ 0 & , \; x \notin [a, b] \end{cases}$$

It can be shown that

- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$
- $\mathbb{E}[X] = \dfrac{a+b}{2}$
- $\mathsf{Var}(X) = \dfrac{(b-a)^2}{12}$



Other important distributions

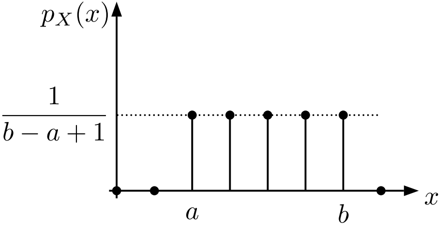
Discrete uniform distribution

X has *discrete uniform distribution* between $a \in \mathbb{N}$ and $b \in \mathbb{N}$, if

$$p_X(x) = \begin{cases} \frac{1}{b-a+1} & , \ a \le x \le b \\ 0 & , \ \text{otherwise} \end{cases}$$

It can be shown that

- $\displaystyle\sum_{k=a}^{b} p_X(k) = 1$
- $\mathbb{E}[X] = \dfrac{a+b}{2}$
- $\mathsf{Var}(X) = \dfrac{(b-a+1)^2 - 1}{12}$



Other important distributions

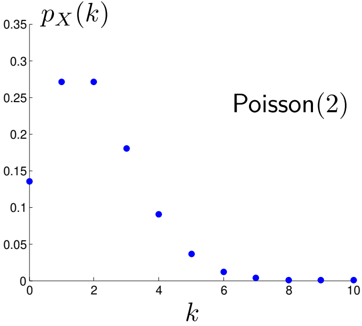
Poisson distribution

X has *Poisson distribution* with parameter λ, $X \sim \text{Poisson}(\lambda)$, if

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k\,!}, \qquad k = 0, 1, \ldots$$

It can be shown that

- $\sum_{k=0}^{+\infty} p_X(k) = 1$
- $\mathbb{E}[X] = \lambda$
- $\text{Var}(X) = \lambda$



Other important distributions

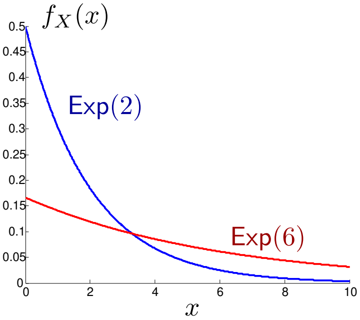
Exponential distribution

X has *exponential distribution* with parameter λ, $X \sim \text{Exp}(\lambda)$, if

$$f_X(x) = \lambda e^{-\lambda x}\,, \qquad \text{for } x \geq 0$$

It can be shown that

- $\displaystyle\int_0^{+\infty} f_X(x)\,dx = 1$
- $\mathbb{E}[X] = \dfrac{1}{\lambda}$
- $\text{Var}(X) = \dfrac{1}{\lambda^2}$



Other important distributions

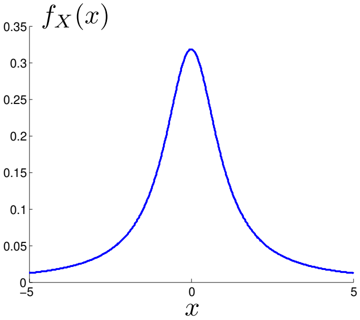
Cauchy distribution

X has *standard Cauchy distribution*, if

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

It can be shown that

- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$
- $\mathbb{E}[X] = \infty$
- $\mathsf{Var}(X) = \infty$



Proofs

Proofs: Properties of the expected value

Proofs just for continuous RVs; for discrete RVs it's similar

$\mathbb{E}\big[a\,X + b\big] = a\,\mathbb{E}[X] + b$

Proof

$$\mathbb{E}\big[a\,X + b\big] = \int (a\,x + b) f_X(x)\,dx = a\underbrace{\int x f_X(x)\,dx}_{\mathbb{E}[X]} + b\underbrace{\int f_X(x)}_{1}\,dx$$

$$= a\,\mathbb{E}[X] + b\,,$$

where we used the linearity of the integral, the definition of expected value, and the fact that $f_X(x)$ is a pdf. $\square$

Proofs: Properties of the expected value

$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Proof

$$
\begin{aligned}
\mathbb{E}[X+Y] &= \int\int (x+y) f_{XY}(x,y)\,dx\,dy \\
&= \int\int x\, f_{XY}(x,y)\,dx\,dy + \int\int y\, f_{XY}(x,y)\,dx\,dy \\
&= \int x \Bigg(\underbrace{\int f_{XY}(x,y)\,dy}_{f_X(x)} \Bigg) dx + \int y \Bigg(\underbrace{\int f_{XY}(x,y)\,dx}_{f_Y(y)} \Bigg) dy \\
&= \int x f_X(x)\,dx + \int y f_Y(y)\,dy \\
&= \mathbb{E}[X] + \mathbb{E}[Y]\,,
\end{aligned}
$$

where we used Fubini's theorem and computed marginal pdfs. □

Proofs: Properties of the expected value

If X and Y are independent, then $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Proof

$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \int \int x\, y\, f_{XY}(x, y)\, dx\, dy \\
&= \int \int x\, y\, f_X(x)\, f_Y(y)\, dx\, dy \\
&= \int y \Bigg(\underbrace{\int x\, f_X(x)\, dx}_{\mathbb{E}[X]} \Bigg) f_Y(y)\, dy \\
&= \mathbb{E}[X] \int y\, f_Y(y)\, dy \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y]\,,
\end{aligned}$$

where we used the independence of X and Y in the 2nd equality, and the fact that $\mathbb{E}[X]$ is a number in the 4th equality. $\square$

Proofs: Properties of the variance

The property $\mathsf{Var}(X) \geq 0$ follows from the definition by observing that the expected value of a nonnegative RV is always nonnegative:

$$\mathsf{Var}(X) = \mathbb{E}\Big[\underbrace{\big(X - \mathbb{E}(X)\big)^2}_{\geq 0}\Big] \geq 0\,.$$

For example, for a continuous RV, if $Y \geq 0$, then

$$\mathbb{E}[Y] = \int \underbrace{y \cdot f_Y(y)}_{\geq 0}\, dy \geq 0\,.$$

$\square$

Proofs: Properties of the variance

$\mathsf{Var}(X) = \mathbb{E}\big[X^2\big] - \big(\mathbb{E}[X]\big)^2$

Proof

$$\begin{aligned}\mathsf{Var}(X) &= \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] \\ &= \mathbb{E}\Big[X^2 - 2X \cdot \mathbb{E}[X] + \big(\mathbb{E}[X]\big)^2\Big] \\ &= \mathbb{E}\big[X^2\big] - 2\mathbb{E}[X] \cdot \mathbb{E}[X] + \big(\mathbb{E}[X]\big)^2 \\ &= \mathbb{E}\big[X^2\big] - \big(\mathbb{E}[X]\big)^2,\end{aligned}$$

where the third equality uses the linearity of the expectation and fact that $\mathbb{E}[X]$ is a constant, i.e., $\mathbb{E}\big[\mathbb{E}[X]\big] = \mathbb{E}[X]$. $\square$

Proofs: Properties of the variance

For any scalars a and b, $\mathrm{Var}(a\,X+b) = a^2\,\mathrm{Var}(X)$.

Proof

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

In the second and fourth equalities, we used the linearity of the expectation. $\square$

Proofs: Properties of the variance

If X and Y are independent, then $\mathsf{Var}(X+Y) = \mathsf{Var}(X) + \mathsf{Var}(Y)$

Proof

$$\begin{aligned}
\mathsf{Var}(X+Y) &= \mathbb{E}\Big[\big(X+Y-\mathbb{E}[X]-\mathbb{E}[Y]\big)^2\Big] \\
&= \mathbb{E}\Big[\big((X-\mathbb{E}[X])+(Y-\mathbb{E}[Y])\big)^2\Big] \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathbb{E}\Big[(X-\mathbb{E}[X])\cdot(Y-\mathbb{E}[Y])\Big] \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathbb{E}\big[(X-\mathbb{E}[X])\big]\cdot\mathbb{E}\big[(Y-\mathbb{E}[Y])\big] \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\big(\mathbb{E}[X]-\mathbb{E}[X]\big)\cdot(\mathbb{E}[Y]-\mathbb{E}[Y]) \\
&= \mathsf{Var}(X) + \mathsf{Var}(Y)\,.
\end{aligned}$$

In the fourth equality we used the fact that if X and Y are independent, so are $X-a$ and $Y-b$, for any constants a and b (see next page). $\square$

Proofs: Properties of the variance

Proof that if X and Y are independent, so are $X - a$ and $Y - b$:

$$\begin{aligned}
\mathbb{E}[(X-a)\cdot(Y-b)] &= \mathbb{E}\big[XY - bX - aY + ab\big] \\
&= \mathbb{E}[XY] - b\cdot\mathbb{E}[X] - a\cdot\mathbb{E}[Y] + ab \\
&= \mathbb{E}[X]\cdot\mathbb{E}[Y] - b\cdot\mathbb{E}[X] - a\cdot\mathbb{E}[Y] + ab \\
&= \mathbb{E}[X]\cdot\big(\mathbb{E}[Y]-b\big) - a\cdot\big(\mathbb{E}[Y]-b\big) \\
&= \big(\mathbb{E}[X]-a\big)\cdot\big(\mathbb{E}[Y]-b\big)\,,
\end{aligned}$$

where in the second equality we used the linearity of the expectation, and in the third equality the fact that X and Y are independent. $\square$