Theorem (Expected value rule) Given a random variable X and a function $g:\mathbb{R}\to\mathbb{R}$, we construct the random variable $Y=g(X)$. Then

$$\sum_y y p_Y(y) = \mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_x g(x)p_X(x).$$

Remark (PMF of $Y=g(X)$) The PMF of $Y=g(X)$ is $p_Y(y) = \sum_{x:g(x)=y} p_X(x)$.

Remark In general $g(\mathbb{E}[X]) \neq \mathbb{E}[g(X)]$. They are equal if $g(x)=ax+b$.

Variance, conditioning on an event, multiple r.v.

Definition (Variance of a random variable) Given a random variable X with $\mu=\mathbb{E}[X]$, its variance is a measure of the spread of the random variable and is defined as

$$\operatorname{Var}(X) \triangleq \mathbb{E}\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 p_X(x).$$

Definition (Standard deviation)

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

Properties (Properties of the variance)

- $\operatorname{Var}(aX) = a^2\operatorname{Var}(X)$, for all $a\in\mathbb{R}$.
- $\operatorname{Var}(X+b) = \operatorname{Var}(X)$, for all $b\in\mathbb{R}$.
- $\operatorname{Var}(aX+b) = a^2\operatorname{Var}(X)$.
- $\operatorname{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Example (Variance of known r.v.)

- If $X\sim\operatorname{Ber}(p)$, then $\operatorname{Var}(X)=p(1-p)$.
- If $X\sim\operatorname{Uni}[a,b]$, then $\operatorname{Var}(X)=\frac{(b-a)(b-a+2)}{12}$.
- If $X\sim\operatorname{Bin}(n,p)$, then $\operatorname{Var}(X)=np(1-p)$.
- If $X\sim\operatorname{Geo}(p)$, then $\operatorname{Var}(X)=\frac{1-p}{p^2}$

Proposition (Conditional PMF and expectation, given an event) Given the event A, with $\mathbb{P}(A)>0$, we have the following

- $p_{X|A}(x)=\mathbb{P}(X=x|A)$.
- If A is a subset of the range of X, then: $$p_{X|A}(x) \triangleq p_{X|\{X\in A\}}(x) = \begin{cases}\frac{1}{\mathbb{P}(A)}p_X(x), & \text{if } x\in A,\\ 0, & \text{otherwise.}\end{cases}$$
- $\sum_x p_{X|A}(x)=1$.
- $\mathbb{E}[X|A]=\sum_x x p_{X|A}(x)$.
- $\mathbb{E}[g(X)|A]=\sum_x g(x)p_{X|A}(x)$.

Proposition (Total expectation rule) Given a partition of disjoint events $A_1,\dots,A_n$ such that $\sum_i \mathbb{P}(A_i)=1$, and $\mathbb{P}(A_i)>0$,

$$\mathbb{E}[X] = \mathbb{P}(A_1)\mathbb{E}[X|A_1]+\cdots+\mathbb{P}(A_n)\mathbb{E}[X|A_n].$$

Definition (Memorylessness of the geometric random variable) When we condition a geometric random variable X on the event $X>n$ we have memorylessness, meaning that the "remaining time" $X-n$, given that $X>n$, is also geometric with the same parameter. Formally,

$$p_{X-n|X>n}(i) = p_X(i).$$

Definition (Joint PMF) The joint PMF of random variables $X_1,X_2,\dots,X_n$ is $p_{X_1,X_2,\dots,X_n}(x_1,\dots,x_n)=\mathbb{P}(X_1=x_1,\dots,X_n=x_n)$.

Properties (Properties of joint PMF)

- $\sum_{x_1}\cdots\sum_{x_n} p_{X_1,\dots,X_n}(x_1,\dots,x_n)=1$.
- $p_{X_1}(x_1)=\sum_{x_2}\cdots\sum_{x_n} p_{X_1,\dots,X_n}(x_1,x_2,\dots,x_n)$.
- $p_{X_2,\dots,X_n}(x_2,\dots,x_n)=\sum_{x_1} p_{X_1,X_2,\dots,X_n}(x_1,x_2,\dots,x_n)$.

Definition (Functions of multiple r.v.) If $Z=g(X_1,\dots,X_n)$, where $g:\mathbb{R}^n\to\mathbb{R}$, then $p_Z(z)=\mathbb{P}(g(X_1,\dots,X_n)=z)$.

Proposition (Expected value rule for multiple r.v.) Given $g:\mathbb{R}^n\to\mathbb{R}$,

$$\mathbb{E}[g(X_1,\dots,X_n)] = \sum_{x_1,\dots,x_n} g(x_1,\dots,x_n)p_{X_1,\dots,X_n}(x_1,\dots,x_n).$$

Properties (Linearity of expectations)

- $\mathbb{E}[aX+b]=a\mathbb{E}[X]+b$.
- $\mathbb{E}[X_1+\cdots+X_n]=\mathbb{E}[X_1]+\cdots+\mathbb{E}[X_n]$.

Conditioning on a random variable, independence

Definition (Conditional PMF given another random variable) Given discrete random variables X,Y and y such that $p_Y(y)>0$ we define

$$p_{X|Y}(x|y) \triangleq \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

Proposition (Multiplication rule) Given jointly discrete random variables X,Y, and whenever the conditional probabilities are defined,

$$p_{X,Y}(x,y) = p_X(x)p_{Y|X}(y|x) = p_Y(y)p_{X|Y}(x|y).$$

Definition (Conditional expectation) Given discrete random variables X,Y and y such that $p_Y(y)>0$ we define

$$\mathbb{E}[X|Y=y] = \sum_x x p_{X|Y}(x|y).$$

Additionally we have

$$\mathbb{E}[g(X)|Y=y] = \sum_x g(x)p_{X|Y}(x|y).$$

Theorem (Total probability and expectation theorems) If $p_Y(y)>0$, then

$$p_X(x) = \sum_y p_Y(y)p_{X|Y}(x|y),$$
$$\mathbb{E}[X] = \sum_y p_Y(y)\mathbb{E}[X|Y=y].$$

Definition (Independence of a random variable and an event) A discrete random variable X and an event A are independent if $\mathbb{P}(X=x \text{ and } A)=p_X(x)\mathbb{P}(A)$, for all x.

Definition (Independence of two random variables) Two discrete random variables X and Y are independent if $p_{X,Y}(x,y)=p_X(x)p_Y(y)$ for all x,y.

Remark (Independence of a collection of random variables) A collection $X_1,X_2,\dots,X_n$ of random variables are independent if

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = p_{X_1}(x_1)\cdots p_{X_n}(x_n),\ \forall x_1,\dots,x_n.$$

Remark (Independence and expectation) In general, $\mathbb{E}[g(X,Y)]\neq g(\mathbb{E}[X],\mathbb{E}[Y])$. An exception is for linear functions: $\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y]$.

Proposition (Expectation of product of independent r.v.) If X and Y are discrete independent random variables,

$$\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y].$$

Remark If X and Y are independent, $\mathbb{E}[g(X)h(Y)]=\mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$.

Proposition (Variance of sum of independent random variables) IF X and Y are discrete independent random variables,

$$\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y).$$

Continuous random variables

PDF, Expectation, Variance, CDF

Definition (Probability density function (PDF)) A probability density function of a r.v. X is a non-negative real valued function f_X that satisfies the following

- $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x=1$.
- $\mathbb{P}(a\le X\le b)=\int_a^b f_X(x)\mathrm{d}x$ for some random variable X.

Definition (Continuous random variable) A random variable X is continuous if its probability law can be described by a PDF f_X.

Remark Continuous random variables satisfy:

- For small $\delta>0$, $\mathbb{P}(a\le X\le a+\delta)\approx f_X(a)\delta$.
- $\mathbb{P}(X=a)=0$, $\forall a\in\mathbb{R}$.

Definition (Expectation of a continuous random variable) The expectation of a continuous random variable is

$$\mathbb{E}[X] \triangleq \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x.$$

assuming $\int_{-\infty}^{\infty}|x|f_X(x)\mathrm{d}x<\infty$.

Properties (Properties of expectation)

- If $X\ge 0$ then $\mathbb{E}[X]\ge 0$.
- If $a\le X\le b$ then $a\le\mathbb{E}[X]\le b$.
- $\mathbb{E}[g(X)]=\int_{-\infty}^{\infty} g(x)f_X(x)\mathrm{d}x$.
- $\mathbb{E}[aX+b]=a\mathbb{E}[X]+b$.

Definition (Variance of a continuous random variable) Given a continuous random variable X with $\mu=\mathbb{E}[X]$, its variance is

$$\operatorname{Var}(X)=\mathbb{E}\left[(X-\mu)^2\right]=\int_{-\infty}^{\infty}(x-\mu)^2 f_X(x)\mathrm{d}x.$$

It has the same properties as the variance of a discrete random variable.

Example (Uniform continuous random variable) A Uniform continuous random variable X between a and b, with $a<b$, $(X\sim\operatorname{Uni}(a,b))$ has PDF

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

We have $\mathbb{E}[X]=\frac{a+b}{2}$ and $\operatorname{Var}(X)=\frac{(b-a)^2}{12}$.