Discrete Random Variables

- A **random variable** is a real-valued function defined on the sample space:

$$X : \Omega \to \mathbb{R}$$

- The notation $\{X = x\}$ denotes an event:

$$\{X = x\} = \{\omega \in \Omega | X(\omega) = x\} \subseteq \Omega$$

- The **probability mass function (PMF)** for the random variable X assigns a probability to each event $\{X = x\}$:

$$p_X(x) = \mathbf{P}(\{X = x\}) = \mathbf{P}\Big(\{\omega \in \Omega | X(\omega) = x\}\Big)$$

PMF Properties

- Let X be a random variable and S a countable subset of the real line
- The axioms of probability hold:
  1. $p_X(x) \geq 0$
  2. $\mathbf{P}(X \in S) = \sum_{x \in S} p_X(x)$
  3. $\sum_x p_X(x) = 1$
- If g is a real-valued function, then $Y = g(X)$ is a random variable:

$$\omega \xrightarrow{X} X(\omega) \xrightarrow{g} g(X(\omega)) = Y(\omega)$$

with PMF

$$p_Y(y) = \sum_{x|g(x)=y} P_X(x)$$

Expectation

Given a random variable X with PMF $p_X(x)$:

- $\mathbf{E}[X] = \sum_x x p_X(x)$
- Given a derived random variable $Y = g(X)$:

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

$$\mathbf{E}[X^n] = \sum_x x^n p_X(x)$$

- **Linearity** of Expectation: $\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$.

Variance

The expected value of a derived random variable $g(X)$ is

$$\mathbf{E}[g(X)] = \sum_x g(x) p_X(x)$$

The variance of X is calculated as

- $var(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \sum_x (x - \mathbf{E}[X])^2 p_X(x)$
- $var(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2$
- $var(aX + b) = a^2 var(X)$

Multiple Random Variables

Let X and Y denote random variables defined on a sample space Ω.

- The **joint PMF** of X and Y is denoted by

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

- The **marginal PMFs** of X and Y are given respectively as

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y)$$

Functions of Multiple Random Variables

Let $Z = g(X, Y)$ be a function of two random variables

- **PMF:**

$$p_Z(z) = \sum_{(x,y)|g(x,y)=z} p_{X,Y}(x, y)$$

- **Expectation:**

$$\mathbf{E}[Z] = \sum_{x,y} g(x, y) p_{X,Y}(x, y)$$

- **Linearity:** Suppose $g(X, Y) = aX + bY + c$.

$$\mathbf{E}[g(X, Y)] = a\mathbf{E}[X] + b\mathbf{E}[Y] + c$$

Conditioned Random Variables

- If A is an event with $\mathbf{P}(A) > 0$, then

$$p_{X|A}(x) = \mathbf{P}\big(\{X = x\}|A\big) = \frac{\mathbf{P}\big(\{X = x\} \cap A\big)}{\mathbf{P}(A)}$$

- If Y is a random variable and $\mathbf{P}_Y(y) > 0$, then

$$p_{X|Y}(x|y) = \frac{\mathbf{P}\big(\{X = x\} \cap \{Y = y\}\big)}{\mathbf{P}\big(\{Y = y\}\big)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

- $p_X(x) = \displaystyle\sum_{i=1}^{n} \mathbf{P}(A_i) p_{X|A_i}(x)$

Conditional Expectation

Let X and Y be random variables on a sample space Ω.

- Given an event A with $\mathbf{P}(A) > 0$

$$\mathbf{E}[X|A] = \sum_x x p_{X|A}(x)$$

- If $P_Y(y) > 0$, then

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

- **Total Expectation Theorem:** Let $A_1, \ldots, A_n$ be a partition of Ω. If $\mathbf{P}(A_i) > 0 \ \forall i$, then

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{P}(A_i)\mathbf{E}[X|A_i]$$

Independence

Let X and Y be random variables defined on Ω and let A be an event with $\mathbf{P}(A) > 0$.

- X is independent of A if either of the following hold:

$$p_{X|A}(x) = p_X(x)$$
$$p_{X,A}(x) = p_X(x)\mathbf{P}(A)$$

  for each x.

- X and Y are independent if, for each x and y, either of the following hold:

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

Independence

If X and Y are independent, then the following hold:

- If g and h are real-valued functions, then $g(X)$ and $h(Y)$ are independent.
- $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$
- $var(X+Y) = var(X) + var(Y)$

Given independent random variables $X_1, \ldots, X_n$,

$$var(X_1 + X_2 + \cdots + X_n) = var(X_1) + var(X_2) + \cdots + var(X_n)$$

Canonical Discrete Distributions

	X	$p_X(k)$	$\mathbf{E}[X]$	$var(X)$
Bernoulli	$\begin{cases} 1 & \text{success} \\ 0 & \text{failure} \end{cases}$	$\begin{cases} p & k=1 \\ 1-p & k=0 \end{cases}$	p	$p(1-p)$
Binomial	Number of successes in n Bernoulli trials	$\binom{n}{k} p^k (1-p)^{n-k}$ $k = 0, 1, \ldots, n$	np	np(1-p)
Geometric	Number of trials until first success	$(1-p)^{k-1} p$ $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Uniform	An integer in the interval [a,b]	$\begin{cases} \frac{1}{b-a+1} & k = a, \ldots, b \\ 0 & \text{otherwise} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)(b-a+2)}{12}$
Poisson	Number of rare events	$\frac{e^{-\lambda}\lambda^k}{k!}$ $k = 0, 1, 2, \ldots$	λ	λ

Probability Density Functions (PDF)

For a continuous RV X with PDF $f_X(x)$ (≥ 0),

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b f_X(x)dx \\ P(x \leq X \leq x+\delta) &\approx f_X(x).\delta \\ P(X \in A) &= \int_A f_X(x)dx \end{aligned}$$

Remarks:

- if X is continuous, $P(X = x) = 0 \quad \forall x$!!
- $f_X(x)$ may take values larger than 1.

Normalization property:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Mean and variance of a continuous RV

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x)dx \\
E[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x)dx \\
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)dx \\
&= E[X^2] - (E[X])^2 \ (\geq 0) \\
E[aX + b] &= aE[X] + b \\
\mathrm{Var}(aX + b) &= a^2 \mathrm{Var}(X)
\end{aligned}
$$

Cumulative Distribution Functions

Definition:

$$F_X(x) = P(X \leq x)$$

monotonically increasing from 0 (at $-\infty$) to 1 (at $+\infty$).

- Continuous RV:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt \quad \text{(continuous)}$$

$$f_X(x) = \frac{dF_X}{dx}(x)$$

Cumulative Distribution Functions

Definition:

$$F_X(x) = P(X \leq x)$$

monotonically increasing from 0 (at $-\infty$) to 1 (at $+\infty$).

- Discrete RV:

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k) \text{ (piecewise constant)}$$

$$p_X(k) = F_X(k) - F_X(k-1) \text{ (height of step at k)}$$

Normal/Gaussian Random Variables

Standard Normal RV: $N(0,1)$:

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\
E[X] &= 0, \qquad \mathrm{Var}(X) = 1
\end{aligned}
$$

General normal RV: $N(\mu, \sigma^2)$:

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \\
E[X] &= \mu, \qquad \mathrm{Var}(X) = \sigma^2
\end{aligned}
$$

Normal/Gaussian Random Variables

- if $Y = aX + b$, then $Y \sim N(a\mu + b,\ a^2\sigma^2)$.
- CDF for standard normal $\phi(.)$ can be read in a table.
- To evaluate CDF of a general standard normal, express it as a function of a standard normal:

$$X \sim N(\mu, \sigma^2) \Leftrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$P(X \leq x) = P\Big(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\Big) = \phi\Big(\frac{x - \mu}{\sigma}\Big)$$

where $\phi(.)$ denotes the CDF of a standard normal.

Derived distributions

Def: PDF of a *function* of a RV X with known PDF: $Y = g(X)$.
Method:

- Get the CDF:

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{x|g(x)\leq y} f_X(x)dx$$

- Differentiate: $f_Y(y) = \frac{dF_Y}{dy}(y)$

Law of iterated expectations

$E[X|Y]$ is a random variable that is a function of Y (the expectation is taken with respect to X).
To compute $E[X|Y]$, first express $E[X|Y = y]$ as a function of y.

Law of iterated expectations:

$$E[X] = E[E[X|Y]]$$

(equality between two real numbers)

Law of conditional variances

$\mathrm{Var}(X|Y)$ is a random variable that is a function of Y (the variance is taken with respect to X).
To compute $\mathrm{Var}(X|Y)$, first express

$$\mathrm{Var}(X|Y=y) = E[(X - E[X|Y=y])^2|Y=y]$$

as a function of y.

Law of conditional variances:

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y])$$

(equality between two real numbers)

Sum of a random number of iid RVs

N discrete RV, X_i i.i.d and independent of N.
$Y = X_1 + \ldots + X_N$. Then:

$$\begin{aligned} E[Y] &= E[X]E[N] \\ \mathrm{Var}(Y) &= E[N]\mathrm{Var}(X) + (E[X])^2\mathrm{Var}(N) \end{aligned}$$

Covariance and Correlation

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

Correlation: (has no dimension)

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad \in [-1, 1]$$

By definition, X, Y are uncorrelated if and only if $\mathrm{Cov}(X, Y) = 0$.
Remark: X, Y independent $\Rightarrow$ $\mathrm{Cov}(X, Y) = 0$
(the converse is not true)

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

Some Continuous Random Variables

	$f_X(x)$	$F_X(x)$	$\mathbf{E}[X]$	$\mathrm{var}(X)$
Uniform ($[a,b]$)	$\begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & \text{o.w.} \end{cases}$	$\begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & \text{o.w.}(x > b) \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential (λ)	$\begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & \text{o.w.} \end{cases}$	$\begin{cases} 1 - e^{-\lambda x}, & x \ge 0 \\ 0, & \text{o.w.} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal (μ, σ^2)	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$		μ	σ^2

Markov Inequality

- If X is a random variable that is nonnegative with probability 1,then

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}, \text{for all } a > 0.$$

- Intuitive meaning - if a nonnegative random variable has a small expectation, then the probability that it takes on a large value must be small.
- There are simple examples which prove that the Markov inequality can be tight. However, in general,the Markov inequality is quite loose .

Chebyshev Inequality

- For any random variable X with finite mean μ and variance σ^2,

$$\mathbf{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}, \text{ for all } k > 0.$$

$$\mathbf{P}(|X - \mu| \geq c\sigma) \leq \frac{1}{c^2}, \text{ for all } c > 0.$$

Chebyshev Inequality

- The Chebyshev inequality is just a special case of the Markov inequality. As such, the bounds it gives are often quite loose.
- Intuitive meaning - a random variable with small variance cannot deviate far from its expected value.
- Common Question - Use Chebyshev inequality to estimate probability a sample average deviates far from its mean.