# Chapter 2 Summary

## Discrete Random Variables

- If $\sum_{x\in\mathbb{R}} \mathsf{P}(X=x) = 1$, then $X$ is a *discrete random variable.*
- The *probability function* is $p_X(x) = \mathsf{P}(X=x)$.
- The *law of total probability*, discrete random variable version, is

$$\mathsf{P}(A) = \sum_{x\in\mathbb{R}} \mathsf{P}(X=x)\,\mathsf{P}(A \mid X=x)\ .$$

- **Degenerate distributions:** $\mathsf{P}(X=c)=1$ for some $c$.
- **Bernoulli distributions:** $X \sim \text{Bernoulli}(\theta)$. Single coin toss. $\mathsf{P}(X=1)=\theta$, $\mathsf{P}(X=0)=1-\theta$ where $0<\theta<1$ is a parameter.
- **Binomial distributions:** $X \sim \text{Binomial}(n,\theta)$. Number of heads in $n$ coin tosses, probability of a head is $\theta$ for each toss where $0<\theta<1$ and $n$ is a positive integer.

$$\mathsf{P}(X=k) = \binom{n}{k}\theta^k(1-\theta)^{n-k} \quad \text{for } k=0,1,\ldots,n\ .$$

  $X = X_1+\cdots+X_n$ where $X_i \sim$ i.i.d. Bernoulli$(\theta)$.

- **Geometric distributions:** $X \sim \text{Geometric}(\theta)$. Number of tails before the first head in repeated coin tosses with head probability $\theta$ for each toss where $0<\theta<1$.

$$\mathsf{P}(X=k) = \theta(1-\theta)^k \quad \text{for } k=0,1,2,\ldots\ .$$

- **Negative Binomial distributions:** $X \sim \text{Negative Binomial}(r,\theta)$. Number of tails before the $r$th head in repeated coin tosses with head probability $\theta$ for each toss where $0<\theta<1$.

$$\mathsf{P}(X=k) = \binom{k+r-1}{r-1}\theta^r(1-\theta)^k \quad \text{for } k=0,1,2,\ldots\ .$$

  $X = X_1+\cdots+X_n$ where $X_i \sim$ i.i.d. Geometric$(\theta)$.

- **Poisson distributions:** $X \sim \text{Poisson}(\lambda)$. For parameter $\lambda>0$,

$$\mathsf{P}(X=k) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \quad \text{for } k=0,1,2,\ldots\ .$$

  If $X_n \sim \text{Binomial}(n,\theta_n)$ for $\lambda = n\theta_n$, $n$ large and $\theta_n$ very small, then $\mathsf{P}(X_n=k) \approx \mathsf{P}(X=k)$.

- **Hypergeometric distributions:** $X \sim \text{Hypergeometric}(N,M,n)$. Sample $n$ balls without replacement from $M$ white and $N-M$ black balls, counting the number of white balls in the sample.

$$\mathsf{P}(X=k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},$$

  where $k = \max\{0, n-N+M\},\ldots,\min\{M,n\}$.

## Absolutely Continuous Random Variables

- If $\sum_{x\in\mathbb{R}} \mathsf{P}(X=x) = 0$, then $X$ is a *continuous random variable.*
- An *absolutely continuous random variable* $X$ has a *density function* $f_X(x)$ such that

$$\mathsf{P}(a<X<b) = \int_a^b f_X(x)\,\mathrm{d}x$$

  for all $a \le b$.

- For density functions, $f(x) \ge 0$ for all $x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$.
- **Uniform distributions:** $X \sim \text{Uniform}(L,R)$.

$$f(x) = 1/(R-L) \quad \text{for } L<x<R\ .$$

  $X \sim \text{Uniform}(0,1)$ is a special case.

- **Exponential distributions:** $X \sim \text{Exponential}(\lambda)$. For parameter $\lambda>0$, the density is

$$f(x) = \lambda\mathrm{e}^{-\lambda x} \quad \text{for } x \ge 0\ .$$

  The right tail probability is $\mathsf{P}(X>x) = \mathrm{e}^{-\lambda x}$.

- **Gamma distributions:** $X \sim \text{Gamma}(\alpha,\lambda)$, for shape parameter $\alpha>0$ and scale parameter $\lambda>0$. The density is

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x} \quad \text{for } x \ge 0,$$

  where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\mathrm{e}^{-x}\,\mathrm{d}x$$

  is the *gamma function*, defined when $\alpha>0$.

  Note that $\Gamma(n+1) = n!$ for nonnegative integer $n$, $\Gamma(1/2) = \sqrt{\pi}$, and $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$.

  Gamma$(1,\lambda)$ is Exponential$(\lambda)$.

  $X = X_1+\cdots+X_n$ where $X_i \sim$ i.i.d. Exponential$(\lambda)$.

- **Normal distributions:** $X \sim \mathrm{N}(\mu,\sigma^2)$. Bell-shaped curves. The density is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-((x-\mu)/\sigma)^2/2} \quad (-\infty<x<\infty)\ .$$

  In the *standard normal distribution*, $\mu=0, \sigma^2=1$ and the density is

$$\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-z^2/2}\ .$$

  If $Z \sim \mathrm{N}(0,1)$, then $X = \mu+\sigma Z \sim \mathrm{N}(\mu,\sigma^2)$.

  The cumulative distribution function of the standard normal distribution is

$$\Phi(z) = \int_{-\infty}^{z} \phi(z)\mathrm{d}z\ .$$

  If $X \sim \mathrm{N}(\mu,\sigma^2)$, then

$$\mathsf{P}(X \le x) = \Phi\left(\frac{X-\mu}{\sigma}\right)\ .$$

- **Beta distributions:** $X \sim \text{Beta}(a, b)$. With parameters $a > 0, b > 0$, the density is

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \quad \text{for } 0 < x < 1 \ .$$

where the *beta function* $B(a,b)$ is defined for $a > 0, b > 0$ as

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} \mathrm{d}x = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \ .$$

Beta(1,1) is Uniform(0,1).

### Cumulative Distribution Functions

- *The cumulative distribution function (cdf)* is defined as $F(x) = \mathsf{P}(X \le x)$.
- $F$ determines the distribution.
- For an absolutely continuous random variable $X$ with density $f$, $\frac{\mathrm{d}}{\mathrm{d}x} F(x) = f(x)$.
- $0 \le F(x) \le 1$ for all $x$, $F$ is nondecreasing.
- $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

### One-Dimensional Change of Variable

- If $X$ is a discrete random variable and $Y = h(X)$, then

$$\mathsf{P}(Y = y) = \sum_{x : h(x) = y} P(X = x) \ .$$

- If $X$ is an absolutely continuous random variable with density $f_X$; if $Y = h(X)$ where $h$ is strictly increasing or strictly decreasing when $f_X(x) > 0$ so that $y = h(x)$ implies that $x = h^{-1}(y)$ is unique; and if $h$ is differentiable with derivative $h'$; then

$$f_Y(y) = f_X(h^{-1}(y)) / |h'(h^{-1}(y))| \ .$$

### Joint Distributions

- The *joint distribution* of random variables $X$ and $Y$ is the collection of probabilities $\mathsf{P}((X, Y) \in B)$ for all measurable subsets $B \in \mathbb{R}^2$.
- The *joint cumulative distribution function* $F_{X,Y}(x, y) = \mathsf{P}(X \le x, Y \le y)$ determines the joint distribution of $X$ and $Y$.
- The joint distribution of $X$ and $Y$ determines the *marginal distributions* of $X$ and of $Y$.
- For discrete random variables $X$ and $Y$, the *joint probability function* $p_{X,Y}(x, y) = \mathsf{P}(X = x, Y = y)$ determines the joint distribution.
- $\sum_x \sum_y p_{X,Y}(x, y) = 1$.
- For absolutely continuous random variables $X$ and $Y$, the *joint density function* $f_{X,Y}(x, y)$ defined so that $\mathsf{P}(a \le X \le b, c \le Y \le d) = \int_c^d \int_a^b f_{X,Y}(x, y) \, \mathrm{d}x \, \mathrm{d}y$ determines the joint distribution.
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1$
- For discrete random variables $X$ and $Y$,

$$\begin{aligned} \mathsf{P}(X = x) &= \sum_y \mathsf{P}(X = x, Y = y) \\ \mathsf{P}(Y = y) &= \sum_x \mathsf{P}(X = x, Y = y) \end{aligned}$$

- For absolutely continuous random variables $X$ and $Y$ with joint density $f_{X,Y}$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}y \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}x \end{aligned}$$

### Conditional Distributions

- For discrete random variables $X$ and $Y$, the *conditional distribution of $Y$ given $X = x$* is the collection of probabilities

$$\mathsf{P}(Y \in B \mid X = x) = \frac{\mathsf{P}(Y \in B \cap X = x)}{\mathsf{P}(X = x)}$$

defined whenever $\mathsf{P}(X = x) > 0$.

- The *conditional probability function* of $Y$ given $X$ is

$$p_{Y|X}(y \mid x) = \frac{\mathsf{P}(X = x, Y = x)}{\mathsf{P}(X = x)}$$

defined whenever $\mathsf{P}(X = x) > 0$.

- For absolutely continuous random variables $X$ and $Y$, the *conditional density of $Y$ given $X = x$* is defined to be

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

defined whenever $f_X(x) > 0$.

- The *conditional distribution* of $Y$ given $X$ is the collection of probabilities

$$\mathsf{P}(Y \in B \mid X = x) = \int_a^b f_{Y|X}(y \mid x) \, \mathrm{d}y$$

for $a \le b$ defined whenever $\mathsf{P}(X = x) > 0$.

### Independence

- Two random variables $X$ and $Y$ are *independent* if $\mathsf{P}(X \in B_1, Y \in B_2) = \mathsf{P}(X \in B_1) \mathsf{P}(Y \in B_2)$ for all measurable sets $B_1, B_2 \subset \mathbb{R}$. In other words, random variables $X$ and $Y$ are independent if any event involving $X$ and any event involving $Y$ are independent.

- Discrete random variables $X$ and $Y$ are independent if and only if $\mathsf{P}(X=x, Y=y) = \mathsf{P}(X=x)\,\mathsf{P}(Y=y)$ for all $x, y \in \mathbb{R}$. In addition, discrete random variables $X$ and $Y$ are independent if and only if $\mathsf{P}(Y=y \mid X=x) = \mathsf{P}(Y=y)$ for all $y \in \mathbb{R}$ and all $x \in \mathbb{R}$ such that $\mathsf{P}(X=x) > 0$.

- Absolutely continuous random variables $X$ and $Y$ are independent if and only if the densities can be chosen so that $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all $x, y \in \mathbb{R}$. In addition, absolutely continuous random variables $X$ and $Y$ are independent if and only if $f_{Y|X}(y \mid x) = f_Y(y)$ for all $y \in \mathbb{R}$ and all $x \in \mathbb{R}$ such that $f_X(x) > 0$.

- If random variables $X$ and $Y$ are independent, then random variables $g(X)$ and $h(Y)$ are also independent.

- Collections of discrete random variables $X_1, X_2, \ldots, X_n$ are independent if

  $$\mathsf{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathsf{P}(X_i = x_i)$$

  for all numbers $x_1, \ldots, x_n \in \mathbb{R}$.

- Collections of absolutely continuous random variables $X_1, X_2, \ldots, X_n$ are independent if density functions can be chosen so that

  $$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

  for all numbers $x_1, \ldots, x_n \in \mathbb{R}$.

- A collection of random variables is *independent and identically distributed*, or *i.i.d.* if they are independent and have the same marginal distributions.

## Order Statistics

- Random variables $X_1, \ldots, X_n$ sorted from smallest to largest are labeled

  $$X_{(1)}, \ldots, X_{(n)}$$

  and are called the *order statistics* of the sample.

- $X_{(1)}$ is the minimum, $X_{(n)}$ is the maximum.

## Multidimensional Change of Variable

- If $h = (h_1, \ldots, h_n) : \mathbb{R}^n \to \mathbb{R}^n$ is a differentiable function, its *Jacobian derivative* is defined as

  $$\frac{\partial(h_1, \ldots, h_n)}{\partial(x_1, \ldots, x_n)} = J(x_1, \ldots, x_n) = \det M$$

  where $M$ is an $n \times n$ matrix whose $i,j$ element is $\frac{\partial h_j}{\partial x_i}$. In particular for $n = 2$,

  $$J = \frac{\partial h_1}{\partial x_1}\frac{\partial h_2}{\partial x_2} - \frac{\partial h_1}{\partial x_2}\frac{\partial h_2}{\partial x_1}$$

- If $X_1, \ldots, X_n$ are absolutely continuous random variables with joint density $f_{X_1,\ldots,X_n}$ and if random variables $Y_i = h_i(X_1, \ldots, X_n)$ are defined so that $h_i$ is differentiable and so that $h = (h_1, \ldots, h_n)$ is one-to-one on the region where $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) > 0$, then

  $$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \frac{f_{X_1,\ldots,X_n}(h^{-1}(y_1, \ldots, y_n))}{|J(h^{-1}(y_1, \ldots, y_n))|}$$

## Convolution

- If $X$ and $Y$ are independent discrete random variables and $Z = X + Y$, then

  $$\mathsf{P}(Z = z) = \sum_x \mathsf{P}(X = x)\,\mathsf{P}(Y = z - x) \quad \text{for all } z \in \mathbb{R}\,.$$

- If $X$ and $Y$ are independent absolutely continuous random variables and $Z = X + Y$, then

  $$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\,\mathrm{d}x \quad \text{for all } \; z \in \mathbb{R}\,.$$

## Simulation

- Some special distributions can be simulated from directly as functions of uniform random variables.

- If $U \sim \text{Uniform}(0,1)$, then $X = (R - L)U + L \sim \text{Uniform}(L, R)$.

- If $U \sim \text{Uniform}(0,1)$, then $X = -\ln(U)/\lambda \sim \text{Exponential}(\lambda)$.

- If $U_1, U_2 \sim$ i.i.d. Uniform$(0,1)$, and we define

  $$X = \sqrt{-2\ln(U_1)}\cos(2\pi U_2)$$

  and

  $$Y = \sqrt{-2\ln(U_1)}\sin(2\pi U_2),$$

  then $X \sim \text{N}(0,1)$, $Y \sim \text{N}(0,1)$, and $X$ and $Y$ are independent.

- If $Z \sim \text{N}(0,1)$, then $X = \mu + \sigma Z \sim \text{N}(\mu, \sigma^2)$.

- If $X$ is a random variable with cumulative distribution function $F$, then $F^{-1}(t) = \min\{x : F(x) \geq t\}$ for $0 < t < 1$ is the *inverse cdf* or *quantile function* of $X$.

- If $U \sim \text{Uniform}(0,1)$, then $Y = F^{-1}(U)$ is a random variable with cdf $F$.